Evaluation of The PreK–12 STEM Pathway

EIR Grant U411C180223

October 2024

Submitted to:
Community Training and Assistance Center (CTAC)
One Boston Place, Suite 2606
Boston, MA 02108

Submitted by:
Abt Global LLC
6130 Executive Boulevard
Rockville, MD 20852

About This Report

This report provides findings from the Abt Global independent evaluation of the PreK–12 STEM Pathway program. The evaluation was funded by the Community Training and Assistance Center (CTAC) Education Innovation and Research (EIR) Early Phase grant (EARLY11; Award No. U411C180223).

The views expressed in this report do not necessarily reflect the views or policies of the U.S. Department of Education.

Abt Global LLC | 6130 Executive Boulevard | Rockville, MD 20852

Contents

List of Exhibits

Acknowledgements

We are grateful for the collaboration and support from multiple individuals at CTAC and Tracy Unified School District as we conducted the evaluation of the PreK–12 STEM Pathway program: Jeffrey Edmison, Guodong Liang, Rob Pecot, Zhaogang Qiao, Dean Reese, Scott Reynolds, Debra Schneider, Bill Slotnik, Brian Stephens, and Julianna Stocking. Thank you for all your efforts to help make the program implementation and the evaluation a success.

We appreciate the technical assistance provided by Jessaca Spybrook and Bruce Randel through the EIR technical assistance team. We are also grateful for the comparative short interrupted time series toolkit provided by Cristofer Price, also on the EIR technical assistance team.

Executive Summary

The Community Training and Assistance Center (CTAC), a nonprofit organization focused on supporting educational innovation and community change, received an Education Innovation and Research (EIR) Early Phase grant from the U.S. Department of Education[1] in 2018 to develop and implement the PreK–12 STEM Pathway intervention in California's Tracy Unified School District (TUSD). Abt Global (Abt) was hired to conduct an independent evaluation of the implementation and impacts of the PreK–12 STEM Pathway between the 2019–20 through 2022–23 school years. We provide this report to document the work conducted under this grant agreement.[2]

PreK–12 STEM Pathway Program

The PreK–12 STEM Pathway is designed to provide all students with equitable access to rigorous, relevant, and engaging science, technology, engineering, and mathematics (STEM) experiences inside and outside the classroom. Teams of teachers, district staff, and outside stakeholders collaborate to create and implement project-based learning STEM curricular units appropriate for grades PreK–12. The intervention is based on three promising strategies: (1) placing engineering and computer science at the center of student learning, (2) implementing STEM as a PreK–12 pathway, and (3) integrating STEM into the core curriculum to serve all students.

The PreK–12 STEM Pathway has **three key components**:

1. Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science.
2. Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites.
3. Provide professional support structures and capacity building to effectively implement the curricular units.

Implementation Findings

TUSD implemented the PreK–12 STEM Pathway program from school year 2019–20 through school year 2023–24. TUSD rolled out implementation over the course of four cohorts:

- Cohort 1 comprised the majority of TUSD's grades PreK–5 and began implementation in 2019–20.
- Cohort 2 comprised the majority of TUSD's grades 6–12 and began implementation in 2021–22.
- Cohort 3 comprised the remainder of TUSD's grades PreK–5 and began implementation in 2022–23.
- Cohort 4 comprised the remainder of TUSD's grades 6–12 and began implementation in 2023–24, after the implementation study concluded data collection.[3]

---

1 EIR grant # U411C180223.

2 The main text and executive summary of this report are intended for a broad audience to understand the design and data associated with the report. The technical appendix is intended to meet grant reporting requirements. As these two sections serve different audiences and purposes, the data may be presented or organized in a slightly different manner in each section, as we find befits each audience. The data and findings remain consistent across the two complementary sections.

3 The scope of this evaluation covers TUSD's 16 non-charter schools serving mainstreamed populations. Twelve of these schools received the PreK–12 STEM Pathway as part of the impact study's treatment condition

To implement this program, TUSD and CTAC developed a multi-tiered system of leadership teams, including instructional, administrative, and teacher leaders from the district, each school, and every grade and STEM content area to develop, revise, and train the districts' teachers on 52 multidisciplinary, hands-on, project-based curricular units for each grade and STEM content area in the district. This network of leadership teams was supplemented and supported by district administration, CTAC, local industry leaders, and national experts in the field.

Disruptions from the global COVID-19 pandemic created an unusually difficult context in which to implement and evaluate the intervention. During the spring of school year 2019–20 and the entire 2020–21 school year, TUSD delivered instruction online. Program leadership teams modified the Pre-K–12 STEM Pathway's grades PreK–5 integrated curricular units for virtual delivery of the program. They also worked with district leaders to prioritize and distribute materials to students at home to provide students with hands-on experiences, though they used a virtual learning environment. Despite CTAC and TUSD's creative approaches (e.g., modifying the intervention to be delivered online) and commitment to continuation with the initiative, these conditions created a context unlikely to be replicated in future implementation.

Impact Findings

This impact study's three confirmatory (or primary) research questions looked at the effects of the PreK–12 STEM Pathway program on English language arts (ELA), math, and science achievement (collectively, "academic achievement"). We asked 17 additional exploratory research questions, on the impacts of the PreK–12 STEM Pathway program on:

- Two college readiness outcomes (pass rates of Advanced Placement tests and completion rates of California's college preparatory requirements) after two years of exposure to the PreK–12 STEM Pathway.
- Academic achievement after three years of exposure to the PreK–12 STEM Pathway.
- Academic achievement after two years for select student subgroups: multilingual learners ("MLL students"), Hispanic students, boys, and girls.

To examine these outcomes, the impact study relied on publicly available data from the California Department of Education from school years 2014–15 through 2022–23. These available outcomes broadly assess achievement and college readiness but should be expected to be less sensitive and well-aligned to the PreK–12 STEM Pathway than would be study-collected measures of, for example, engineering or computer science.

To estimate the impact of PreK–12 STEM Pathway, we compared outcomes for students receiving the program to those for students in similar schools during the same school year that did not receive the program. [4] Comparison schools were located both in TUSD and in other districts and did not receive any

---

(Cohorts 1 and 2), and four were included in the impact study's comparison condition. Grades PreK–5 at another two schools (Cohort 3) received the PreK–12 STEM Pathway beginning in the last year of implementation data collection and are therefore also included in the implementation study. The remaining grades 6–12 from the comparison condition schools (Cohort 4) received the PreK–12 STEM Pathway after this study concluded data collection and are not included in the implementation study. TUSD also includes an alternative education program, a charter school, and an adult school that are outside the scope of the study.

[4] The grant team designed a staggered rollout over four school years to serve as many TUSD students for as many years as possible, while also allowing some district schools to remain in the comparison condition within that school year. This design—in particular the inclusion of some TUSD schools in the comparison condition—

PreK–12 STEM Pathway programmatic elements during the study. Comparison schools were identified through a matching process, which was conducted separately for each of the 20 analyses to provide the strongest possible comparison group for each analysis. We used a quasi-experimental design known as a comparative short interrupted time-series (C-SITS) design to estimate the treatment effects.

The study did not find any statistically significant effects of the PreK–12 STEM Pathway on achievement or college readiness outcomes.

Of the 12 subgroup analyses run on academic achievement, one—the impact on MLL students in science—was positive and statistically significant. Given that we estimated 20 individual treatment effects, we would expect at least one to be significant by chance; best practice in the field therefore suggests interpreting this finding as preliminary and worthy of follow-up studies to better understand the programmatic effects and its variation within and between subgroups of students (Westfall, Tobias, & Wolfinger, 2011; Chen, Feng, & Yi, 2017; Schochet, 2008).

This impact analysis was limited by the COVID-19 pandemic in two ways. First, C-SITS designs rely on an assumption that baseline trends would continue in both treatment and comparison, absent the intervention (Hallberg et al., 2018); this assumption does not hold, given the disruption caused by the COVID-19 pandemic. Additionally, many districts opted not to administer state tests during the 2020–21 school year (ETS, 2023a; 2023b); the resulting large rates of data missingness for academic achievement substantially limited the pool of eligible comparison schools for academic achievement outcomes.

Conclusion

CTAC and TUSD designed the PreK–12 STEM Pathway to provide all students with equitable access to project-based STEM learning opportunities. Through this grant, TUSD and CTAC established a multi-tiered system of work teams, created 52 integrated, PreK–12 curricular units, and provided professional development, structures, and capacity building for teachers and administrators. The PreK–12 STEM Pathway aimed to provide rich, high-quality STEM experiences for students to increase student achievement on state tests, increase Advanced Placement exam pass rates, and increase completion of college preparatory courses.

A variety of factors could have contributed to the study not detecting effects of the PreK–12 STEM Pathway for 19 of the 20 outcomes estimated, including methodological challenges brought on by the COVID-19 pandemic. Future work that looks to explore preliminary positive impact estimates of the PreK–12 STEM Pathway on MLL students in science achievement may consider looking to the integrated and project-based nature of the curricular units developed within this grant.

Despite the challenges that arose during implementation, CTAC and TUSD were able to build on their existing strong partnership and draw on superintendent support for the PreK–12 STEM Pathway to engage district leaders and encourage school and teacher buy-in. The program was redesigned to help teachers pivot to teaching the PreK–5 STEM curriculum virtually during the COVID-19 public health emergency, including getting materials to students so they could fully participate in hands-on learning. By doing so, the PreK–12 STEM Pathway provided access to computer science and engineering curricula to all teachers to use with their students. Lessons learned from this study will inform continued efforts to develop and implement high-quality, hands-on STEM experiences for students.

---

allows the study to meet What Works Clearinghouse standards, as required by the grant funding the program's development and implementation.

1. The PreK–12 STEM Pathway Overview

Integrating science, technology, engineering, and mathematics (STEM) instruction is a growing trend in PreK–12 education, likely in response to a growing demand for these skills in the workforce. In 2021, 36.8 million people were employed in STEM occupations in the United States, meaning that STEM-related occupations accounted for 24% of all employment. However, women and Hispanic/Latinx, Black, American Indian, and Alaska Native people are underrepresented in STEM occupations (National Science Board, 2024). A 2017 report by the Bureau of Labor Statistics identified engineering and computer science as the largest occupation groups in STEM and the STEM occupation groups with the greatest prospect of future growth (Fayer et al., 2017). Though the need for a STEM workforce grows, elementary, middle, and secondary students' mathematics scores have declined from 2019 to 2022, likely related, at least in part, to disruptions to instruction nationwide during the COVID-19 pandemic (National Science Board, 2024).

Historically, STEM curricula have comprised disconnected courses in math and science with little focus on engineering and technology. However, the importance of providing authentic, integrated STEM activities has become a larger conversation for schools and districts in recent years (National Academies of Sciences, Engineering, and Medicine, 2019). There is a growing movement to implement programming that integrates STEM into the core curriculum for students in grades PreK–12 (Furner & Kumar, 2007; Roehrig et al., 2021). According to Roehrig et al.'s (2021) framework, key characteristics of integrated STEM include a focus on real-world problems, engagement in engineering design, context integration, content integration, engagement in authentic STEM practices, 21st century skills, and informing students about STEM careers.

Some research has shown that integrating STEM concepts has positive effects on students' understanding of science concepts (Anwar et al., 2022; Guzey et al., 2016; Wendell & Rogers, 2013; Cunningham et al., 2020), but others have found no effect on student learning (e.g., Guzey et al., 2017). Because there are different ways of integrating STEM concepts into curricula, it could be that different ways of integrating STEM concepts could have different effects on student learning. Given the importance of STEM-related fields in today's economy, research into how STEM curricula can affect student learning and engagement can be particularly useful for schools and districts as they work to increase student learning and engagement in STEM disciplines and prepare students for future STEM careers.

Engineering and computer science curricula allow students to apply their math and science knowledge to real-world situations. One study found that elementary school students participating in an engineering curriculum that emphasized student engagement and the connection between engineering and science learning outperformed students participating in traditional engineering content on measures of engineering and science learning (Cunningham et al., 2020). Another found that delivering an engineering curriculum in grade 8 lessened the achievement gap in science for Black and Hispanic students (Cantrell et al., 2006). Additionally, middle school students who received science curricula focused on project-based learning reported significantly higher engagement and confidence in science and more positive attitudes toward it (Basche et al., 2016) and scored higher on science assessments (Harris et al., 2015) than did students who did not participate in project-based learning. Emphasizing engineering and computer science in STEM instruction allows students to engage in hands-on learning that builds interest in STEM and prepares students for future STEM careers.

Building teacher confidence and knowledge in delivering STEM curriculum is important for ensuring a rich STEM experience for students. A recent systematic review of STEM teacher professional development by Huang et al. (2022) found that most professional development centered on how teachers could best teach STEM content in the classroom. Typical approaches to professional development included learning by design, scaffolding authentic experiences, collaborating with peers, and reflecting on practices. However, most studies of STEM professional development for teachers did not include a

control group, and many did not explore changes in teacher knowledge or behavior. More rigorous research into how teacher professional development in STEM affects teacher practices and student learning could be beneficial for schools and districts.

In this study, the Community Training and Assistance Center (CTAC), a nonprofit organization focused on supporting educational innovation and community change, received an Education Innovation and Research (EIR) Early Phase grant from the U.S. Department of Education[5] in 2018 to develop and implement the PreK–12 STEM Pathway intervention in California's Tracy Unified School District (TUSD).

The PreK–12 STEM Pathway is designed to provide all students with equitable access to rigorous, relevant, and engaging STEM experiences inside and outside the classroom. A key aspect of the PreK–12 STEM Pathway is collaboration of teams of teachers, district staff, and outside stakeholders to create and implement project-based, integrated STEM curricular units appropriate for grades PreK–12. The intervention promotes quality STEM instruction through three main avenues: (1) placing engineering and computer science at the center of student learning; (2) implementing STEM as a PreK–12 "pathway," meaning a connected set of STEM curricula and learning experiences that students experience throughout their education from PreK through grade 12; and (3) integrating STEM into the core curriculum to serve all students.

1.1 Program Components

The PreK–12 STEM Pathway intervention comprises three key components: build project-based STEM curricular units, establish a multi-tiered system of leadership teams, and provide professional support structures and capacity building to support school and classroom level implementation of the curricular units (see Logic model, Exhibit 1-2). We describe each in turn:

1. **Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science.** Teams of teachers formed two Standards and Curriculum Teams to develop and revise the curricular units. One team developed the curricular units for grades PreK–5, the other for grades 6–12. TUSD trained the members of the Standards and Curriculum Teams. The trained teachers then collaboratively developed the curricular units (Exhibit 1-1); they also were responsible for revising the curricular units using teacher feedback after initial implementation.

   Exhibit 1-1. PreK–12 STEM Pathway Curricular Units

Grade Level/Subject	Number of Curricular Units
PreK, K, 1, 2, 3, 4, 5	4 curricular units per grade
6, 7, 8	2 science curricular units and 2 math curricular units per grade
Biology, Chemistry, Physics, Geometry, Algebra I, Algebra II	2 curricular units per subject

   The interdisciplinary curricular units integrated grade-level California state STEM standards, state English language arts (ELA) standards, and hands-on learning experiences. The state STEM standards included the computer science standards California adopted in the 2018–19 school year (California State Board of Education, 2022). For example, one of the grade 2 curricular units, *Disperse the Seeds*, focuses on agricultural engineering. Students create a device that mimics the way seeds are dispersed in nature and create a blueprint of their device using Microsoft Paint 3D. The curricular unit incorporates science standards through developing a model that mimics the function of an animal in dispersing seeds or pollinating plants and math standards through estimating length in

[5] EIR grant # U411C180223

inches, feet, centimeters, and meters. One of the grade 8 curricular units, *Create a Cable*, focuses on materials engineering. Students are challenged to build a way to redirect light to enable effective long-distance communication, including programming micro:bit to communicate a transmission to another micro:bit. The focal science standard is integrating qualitative scientific and technical information to support the claim that digitized signals are a more reliable way to encode and transmit information than analog signals. The related math concepts are calculating and graphing data based on motion tests and using the distance-time equation for speed.

TUSD teachers were expected to deliver the curricular units during the school year. During the first two years of the unit's implementation, the Standards and Curriculum Teams collected teacher end-of-year feedback to support refinement of the curricular units.

Field-based learning experiences were complementary to the PreK–12 STEM curricular units to further expand the frequency and enhance the quality of STEM-related experiences.

2. **Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites.** The PreK–12 STEM Pathway included five types of work teams:

   - The Community Collaboratory, a diverse set of educational, community, and STEM industry leaders, aligned the curriculum to industry standards and provided field-based learning experiences for students.
   - Two District Implementation Teams (DITs), consisting of district administrators and teacher leaders, supported implementation of the PreK–5 and 6–12 curriculum, respectively, within TUSD.
   - Site Implementation Teams (SITs) were established at each participating school and included a school administrator and teachers, who supported implementation of the curriculum within the school.
   - The Technical Working Group (TWG), comprising experts from STEM-focused industries and PreK–12 education, advised the Leadership Council on developments in STEM that could improve program implementation.
   - The Leadership Council, which included the CEO of CTAC and the Superintendent of TUSD, oversaw and directed program implementation.

3. **Provide professional support structures and capacity building to effectively implement the curricular units.** CTAC and TUSD provided support to effectively deliver the PreK–12 STEM curricular units created by the Standards and Curriculum Teams. Support included:

   - Development and implementation of site plans at each school to address the project-based learning requirements of the curriculum.
   - Training for DIT and site administrators.
   - Assistance for SITs throughout the school year from district and program staff, such as planning days and training.
   - High-level training on curriculum topics for teachers delivering the curriculum.
   - Executive leadership investment, including individual meetings throughout each year with site leaders focused on STEM implementation.

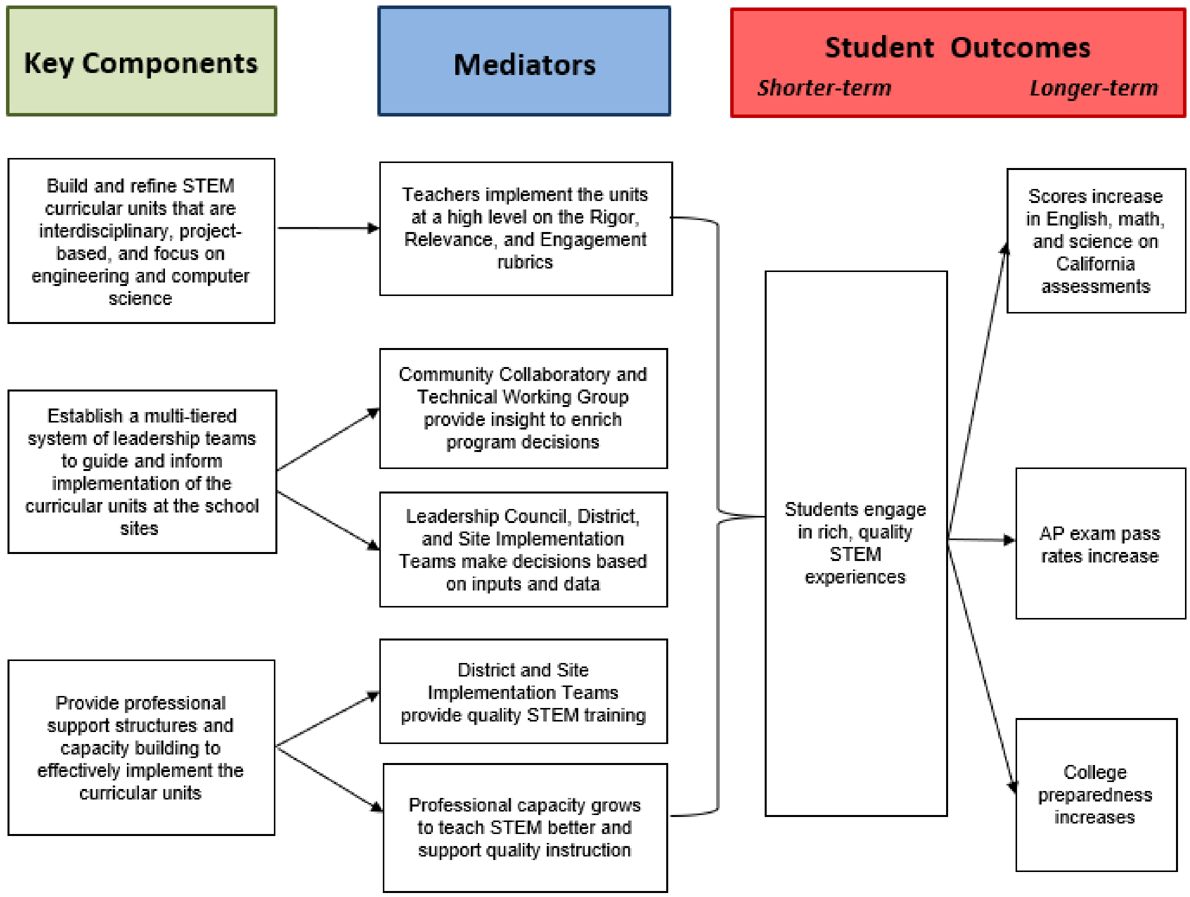
These three key components were designed to strengthen content development along a PreK–12 STEM pathway in support of PreK–12 STEM learning and teaching. The curricular units provided a dual focus

on increasing academically rigorous opportunities for all students and teachers while also building a strong understanding of STEM career options. The components also provided professional support structures and capacity building throughout all levels of the district (e.g., district and school leaders as well as classroom teachers) on instructional leadership, interdisciplinary STEM instruction, and project-based learning.

PreK–12 STEM Pathway aims to provide every student with equitable access to a rigorous STEM learning trajectory from PreK through grade 12, as well as effective learning experiences within students' STEM courses. The PreK–12 STEM Pathway logic model is built on the idea that effective learning experiences are the result of high-quality planning, staffing, and support from a broad array of stakeholders who work together to continuously improve teacher practice and student learning. Through the creation of the PreK–12 STEM curricular units and increased professional support structures, the model hypothesizes that teachers will implement STEM curricular units at a high level of fidelity and receive support from DITs and SITs to grow their professional capacity to teach STEM. Under this model, the various work teams will ensure that the STEM curriculum is informed both by outside experts and by those implementing the curriculum.

Additionally, the model hypothesizes that involving district and site leaders in the decision-making related to support and staffing will increase understanding of, effectiveness in, and accountability for implementing the STEM curriculum and, ultimately, higher-quality STEM instruction. Moreover, the increased support from district executive leadership, district staff, and site staff will also increase administrator capacity to support quality instruction. Under this model, these mediating factors (administrator capacity and teacher implementation) will lead to rich, high-quality STEM experiences for students, which will result in increased student achievement on state tests, increased Advanced Placement (AP) exam pass rates, and increased college preparedness.

Exhibit 1-2. The PreK–12 STEM Pathway Logic Model



Key: AP=Advanced Placement. STEM=science, technology, engineering, mathematics.

1.2 Implementation Context

Tracy Unified School District, serving 13,925 students, is located in San Joaquin County in Northern California's Central Valley (National Center for Education Statistics, n.d.-a). The county, with a population of 779,233, has experienced rapid growth due to overflow from the nearby San Francisco Bay Area (U.S. Census Bureau, n.d.).

The district's demographics reflect the area's diversity. Hispanic or Latino students comprise 58% of the district; non-Hispanic or Latino students comprise the other 42%. Among students who are not Hispanic or Latino, the district comprises White students (19%), Asian students (14%), students of two or more races (5%), and Black or African American students (4%; National Center for Education Statistics, n.d.). This diverse makeup mirrors broader trends in San Joaquin County and the Central Valley region (Public Policy Institute of California, 2021; U.S. Census Bureau, n.d.).

Linguistic diversity is a notable feature of the district, with 40% of students speaking a language other than English at home (National Center for Education Statistics, n.d.). Spanish is the most common non-English language (spoken by 29% of students in the district), followed by Asian and Pacific Islander Languages (3%; National Center for Education Statistics, n.d.).

Tracy's median household income of $105,881 exceeds both state ($91,551) and national ($75,149) averages, as well as the county average ($86,056), indicating a relatively affluent community within the

broader region (National Center for Education Statistics, n.d.-b; U.S. Census Bureau, n.d.). The district also exhibits strong digital connectivity, with 92.2% of households having broadband internet access (National Center for Education Statistics, n.d.-b). This technological infrastructure may be bolstered by the presence of major employers like Amazon and Tesla and reflects the county's growing technology sector and proximity to the tech-centric Bay Area (San Joaquin County, n.d.).

The COVID-19 pandemic created an unusually and unfortunately difficult context for this project; given that this project was funded for the initial development and testing of this PreK–12 STEM Pathway intervention, best practice would have been to test the intervention in "ideal" circumstances (Institute of Education Sciences, 2013). CTAC and TUSD received grant funding for this project during the 2018–19 school year, which served as a planning and development year for the project. The first year of program implementation, for grades PreK–5, was the 2019–20 school year, when the COVID-19 pandemic disrupted learning across the United States, closing schools and resulting in sharp declines in student academic performance. TUSD transitioned to online learning in spring 2020 and resumed in-person learning in fall 2021. CTAC and TUSD rolled the program out to grades 6–12 beginning in the 2021–22 school year.

In California, from the 2019 pre-pandemic testing to the 2022 school year (the first post-pandemic year in which states did not receive a waiver to allow districts to opt out of ESSA-required standardized test requirements), the percentage of students meeting or exceeding state standards dropped for all racial subgroups in ELA and Math (EdSource, 2023). In 2023, the percentage of students in California meeting or exceeding state standards on English and Math is still below pre-pandemic levels. For multilingual learners (MLLs), the percentage meeting or exceeding ELA standards in California largely remains below pre-pandemic levels. For grade 3 MLLs, those meeting or exceeding ELA standards are at the lowest level recorded (16%). In science, across all grades, just 2% of MLLs met or exceeded science standards in 2023 (Buenrostro, 2024). This pandemic-related context was obviously a difficult time for everyone working at or served by schools.

2. Evaluation Overview

Abt Global (Abt) served as the independent external evaluator for the PreK–12 STEM Pathway. Abt's evaluation included a **fidelity of implementation study** and a **study of impacts on student outcomes**.

The implementation study focused on the key components of PreK–12 STEM Pathway during each of the four school years 2019–20, 2020–21, 2021–22, and 2022–23. As described in Section 1, these key components included building and refining STEM curricular units, establishing a multi-tiered system of leadership teams, and providing professional support structures and capacity building to effectively implement the curricular units. Teacher implementation of the curricular units was conceived of as a mediating factor and was out of scope for this study.

The impact study, described in Section 4, used a quasi-experimental C-SITS design to compare outcomes for the "treatment" schools that implemented the PreK–12 STEM Pathway to outcomes for a matched group of "comparison" schools that did not implement it during the study period. The impact study assessed outcomes in four domains: English language arts (ELA) achievement, math achievement, science achievement, and college readiness.

2.1 Pre-Registration of the Study Plan

Abt pre-registered confirmatory analyses of the PreK–12 STEM impact study plan in the Registry of Efficacy and Effectiveness Studies (REES) before collection of any outcome data (Caswell, 2021, Registry ID: 6420.1).

The study team conducted all analyses described in the pre-registered impact study plan. However, it made a few revisions to its plans for those analyses following the REES entry. The study team revised its plan to reflect the study's updated sample sizes after matching treatment schools with comparison schools. The study team also updated its impact model based on updated guidance on best practices for using C-SITS models. The study team added exploratory analyses to the main analyses (the study's confirmatory contrasts) submitted in the study's REES entry.

The updated design plan is described in greater depth in Section 4.

2.2 Research Questions

The PreK–12 STEM Pathway impact study asked both confirmatory research questions that were pre-registered in REES before outcome data collection began and exploratory research questions that were not pre-registered. **Confirmatory research questions** signal the main outcomes of interest for a study. They typically align with the goals of the program being evaluated and pertain to outcomes the program is most likely to affect. Sometimes the selection of outcomes also reflects sample size considerations (e.g., selecting an outcome measured at one point rather than another point if the former has a larger sample).

This impact study's confirmatory research questions focus on impacts of the program on students after two years of exposure as compared to schools not implementing the PreK–12 STEM Pathway (Exhibit 2-1). These analyses look at the effects across all students with outcome data—as opposed to a specific subgroup of students—across all grades.

Exhibit 2-1. Confirmatory Research Questions

Academic Achievement After Two Years (All Students)
C1. What is the effect of the PreK–12 STEM Pathway on the school mean English language arts achievement of students in grades 3–8 and 11 compared to students in schools not implementing the program after two years of exposure?
C2. What is the effect of the PreK–12 STEM Pathway on the school mean math achievement of students in grades 3–8 and 11 compared to students in schools not implementing the program after two years of exposure?
C3. What is the effect of the PreK–12 STEM Pathway on the school mean science achievement of students in grades 5, 8, and 10–12 compared to students in schools not implementing the program after two years of exposure?

In addition, the impact study addresses several exploratory questions. In general, **exploratory research questions** often involve outcomes that are of secondary interest to the study, that are less directly aligned with the goals of the program, or that cannot be as precisely measured (for example, because they involve smaller samples) as confirmatory outcomes can (Columbia University Irving Institute of Medicine, n.d.). In addition, they might examine outcomes for different subpopulations or other time points than do the confirmatory questions. This study includes exploratory questions that focus on effects after three years of exposure for elementary school students (for Cohort 1 only) and effects for secondary school students on two college readiness outcomes after two years of exposure, as compared to schools not implementing the PreK–12 STEM Pathway (Exhibit 2-2). This impact study considered two college readiness outcomes: Advanced Placement (AP) course pass rates and rates of students who satisfy the University of California (UC) and California State University (CSU) systems' requirements to be considered college prepared.[6]

Exhibit 2-2. Exploratory Research Questions for All Students

Academic Achievement After Three Years (All Cohort 1 Students)
E1. What is the effect of the PreK–12 STEM Pathway on the school mean English language arts achievement of students in grades 3–5 compared to students in schools not implementing the program after three years of exposure?
E2. What is the effect of the PreK–12 STEM Pathway on the school mean math achievement of students in grades 3–5 compared to students in schools not implementing the program after three years of exposure?
E3. What is the effect of the PreK–12 STEM Pathway on the school mean science achievement of students in grade 5 compared to students in schools not implementing the program after three years of exposure?
College Readiness After Two Years of Implementation (All Students)
E4. What is the effect of the PreK–12 STEM Pathway on the rates of students who earn a passing score on at least two Advanced Placement courses compared to students in schools not implementing the program after two years of exposure?
E5. What is the effect of the PreK–12 STEM Pathway on the percentage of students in high school who are classified as "prepared" by the University of California or California State University system compared to students in schools not implementing the program after two years of exposure?

6 The UC and CSU systems have established a uniform minimum set of courses required for admission, which are referred to as the "A-G course requirements" (California Department of Education, 2024c; 2023a). Students must have, through coursework or exam: A=2 years of history, B=4 years of English, C=3 years of mathematics, D=2 years of science, E=2 years of language other than English, F=1 year of visual and performing arts, G=1 year of a college-preparatory elective. Students are classified as "prepared" when they meet these requirements with a grade of C minus or higher and meet at least one additional college readiness criterion. These additional criteria include earning a score of at least a Level 3 on the ELA or math Smarter Balanced Summative Assessment and at least a Level 2 on the other assessment; completing one semester (or the equivalent) of college credit courses with at least a grade of C minus and receiving college credits for each course; earning a score of 3 on one AP exam or score of 4 on one International Baccalaureate exam; or completing the Career and Technical Education Pathway (California Department of Education, 2023a).

Finally, the impact study examines the program's effects after two years of exposure for a few specific subgroups of students compared to similar students in comparison schools that did not implement PreK–12 STEM Pathway: students who have ever been classified as English learners (multilingual learners, or "MLL students"),[7] Hispanic students, boys, and girls (Exhibit 2-3). These subpopulations were identified based on their interest to CTAC, as well as being subpopulations TUSD frequently serves and whose needs CTAC attempted to be responsive to when designing the PreK–12 STEM Pathway. For example, in 2018–22, 58% of students in TUSD were Hispanic/Latino, and 40% spoke a language other than English at home (National Center for Education Statistics, n.d.).

Exhibit 2-3. Exploratory Research Questions for Student Subgroups

Academic Achievement After Two Years (Multilingual Learners)

E6. What is the effect of the PreK–12 STEM Pathway on the school mean English language arts achievement of MLL students in grades 3–8 and 11 compared to MLL students in schools not implementing the program after two years of exposure?

E7. What is the effect of the PreK–12 STEM Pathway on the school mean math achievement of MLL students in grades 3–8 and 11 compared to MLL students in schools not implementing the program after two years of exposure?

E8. What is the effect of the PreK–12 STEM Pathway on the school mean science achievement of MLL students in grades 5, 8, and 10–12 compared to MLL students in schools not implementing the program after two years of exposure?

Academic Achievement After Two Years (Hispanic Students)

E9. What is the effect of the PreK–12 STEM Pathway on the school mean English language arts achievement of Hispanic students in grades 3–8 and 11 compared to students in schools not implementing the program after two years of exposure?

E10. What is the effect of the PreK–12 STEM Pathway on the school mean math achievement of Hispanic students in grades 3–8 and 11 compared to Hispanic students in schools not implementing the program after two years of exposure?

E11. What is the effect of the PreK–12 STEM Pathway on the school mean science achievement of Hispanic students in grades 5, 8, and 10–12 compared to Hispanic students in schools not implementing the program after two years of exposure?

Academic Achievement After Two Years (Boys)

E12. What is the effect of the PreK–12 STEM Pathway on the school mean English language arts achievement of boys in grades 3–8 and 11 compared to boys in schools not implementing the program after two years of exposure?

E13. What is the effect of the PreK–12 STEM Pathway on the school mean math achievement of boys in grades 3–8 and 11 compared to boys in schools not implementing the program after two years of exposure?

E14. What is the effect of the PreK–12 STEM Pathway on the school mean science achievement of boys in grades 5, 8, and 10–12 compared to boys in schools not implementing the program after two years of exposure?

[7] We recognize that "multilingual learner" is a broad category that includes students whose initial language(s) may have been or included English. Additionally, some MLL students whose primary home language is not English may have, at no point, been identified by the district as requiring English learner (EL) services. For this report, we adopt the term with the intention of being inclusive and asset-oriented. The data available to us included the 12 categories related to English learner status for which California reports achievement data. These categories included, but were not limited to, initial fluent English proficient, reclassified fluent English proficient, long-term English learner, at risk of becoming a long-term English learner, and ever-English learner (California Department of Education, n.d.). English learner status is different from other student characteristics in that students are continuously entering and exiting English learner classification. Given that we only had access to publicly available, school-level data, we focus on students who had, at any point, been classified as English learners, creating a more consistent group of students.

Academic Achievement After Two Years (Girls)

E15. What is the effect of the PreK–12 STEM Pathway on the school mean English language arts achievement of girls in grades 3–8 and 11 compared to girls in schools not implementing the program after two years of implementation?

E16. What is the effect of the PreK–12 STEM Pathway on the school mean math achievement of girls in grades 3–8 and 11 compared to girls in schools not implementing the program after two years of implementation?

E17. What is the effect of the PreK–12 STEM Pathway on the school mean science achievement of girls in grades 5, 8, and 10–12 compared to girls in schools not implementing the program after two years of implementation?

3. Implementation Study

This section describes the implementation of the PreK–12 STEM Pathway program in TUSD during the evaluation period. We note that, for this study, we relied solely on implementation data provided by CTAC. These data were at the school and district leadership levels and did not measure teacher- or classroom-level participation. This report, therefore, focuses on implementation primarily at the leadership level.

Key Implementation Study Finding

To implement this program, TUSD and CTAC developed **a multi-tiered system of teams** composed of district teachers and administrators to develop, revise, train, and support the districts' teachers on **52 multidisciplinary, hands-on, project-based curricular units for each grade and STEM content area in the district**. This network of leadership teams was overseen, supplemented, and supported by district administration, CTAC, local industry leaders, and national experts in the field.

During the COVID-19 pandemic, TUSD teams modified the Pre-K–12 STEM Pathway curricular units for **virtual delivery of the program** and worked with district leaders to prioritize and distribute materials to students to maintain a hands-on experience for students.

3.1 Implementation Overview

TUSD and CTAC began development of the PreK–12 STEM Pathway in 2018–19. They supported schools to implement the program beginning in school year 2019–20, rolling out the program in cohorts and eventually serving all TUSD mainstream schools and PreK–12 grades.[8] This implementation study spans the first three of four implementation cohorts, covering school years 2018–19, with the initial formation of work teams, through 2022–23. Between 2019–20 and 2022–23, 14 schools participated in the intervention, with some schools implementing different grade bands in different cohorts:

Cohort 1 included the majority of TUSD's grades PreK–5. It included six of the seven PreK–5 schools and the PreK–5 classrooms in three of the four PreK–8 schools. These Cohort 1 schools contribute four years of implementation study data (2019–20, 2020–21, 2021–22, and 2022–23).

Cohort 2 included the majority of TUSD's grades 6–12. It includes grade 6–8 classrooms in the three aforementioned PreK–8 schools, one middle school, and two high schools. These Cohort 2 schools contribute two years of implementation study data (school years 2021–22 and 2022–23).

Cohort 3 included the remaining TUSD PreK–5 grades. It includes the seventh PreK–5 school and the elementary-grade classrooms in the fourth PreK–8 school. These Cohort 3 schools began implementing the PreK–5 STEM Pathway curricular units in school year 2022–23 and contributed implementation study data for that one school year.

Cohort 4 included the remaining TUSD grades 6–12. It includes grades 6–8 at the PreK–8 school whose elementary grades began implementation in Cohort 3, as well as a middle and high school. Cohort 4 began implementation in 2023–24, and therefore does not contribute data to this implementation study.

8 TUSD also includes an alternative education program, a charter school, and an adult school that are outside the scope of this study.

In this section, we describe implementation throughout the grant period and provide a table of the implementation study cohorts (see Exhibit 3-1).

Exhibit 3-1. Implementation Study Cohorts

Cohort	School Grade Configuration	Implementation School Years
Cohort 1 (PreK–5 STEM)	6 PreK–5 schools 3 PreK–8 schools (only grades PreK–5 received the program)	2019–20 2020–21 2021–22 2022-23
Cohort 2 (6–12 STEM)	3 PreK–8 schools (grades 6–8 began to receive the program) 1 middle school (grades 6–8) 2 high schools (grades 9–12)	2021–22 2022–23
Cohort 3 (PreK–5 STEM)	1 PreK–5 school 1 PreK–8 school (only grades PreK–5 received the program)	2022–23
Total	**14 schools**	**4 years**

Notes: The three PreK–8 schools in Cohort 1 and Cohort 2 are the same schools. Cohort 3 schools' grades 6–8 received the program as part of Cohort 4, beginning in the 2023–24 school year, which is outside the scope of this study. TUSD also includes an alternative education program, a charter school, and an adult school that are outside the scope of the study.

Throughout the four years of implementation, CTAC and TUSD implemented three key activities:

- **Built and refined STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science.**
  - The PreK–5 Standards and Curriculum Team created four STEM curricular units per grade level, resulting in 28 total curricular units. Example grade 3 curricular units included "Move That Toy," which explores how forces affect the world around us; "Growing Up," in which students learn how living things change and grow; "Survive and Thrive," which focuses on how the environment affects living organisms; and "Surviving Tracy Winds," which explores how weather affects the Tracy community. In late May and early June 2020, this team adjusted the curricular units in anticipation of both modified in-class and online implementation during the COVID-19 public health emergency. At the end of each of the first two school years that the curricular units were delivered to students, the team met for "Design Team Days" to adjust the curricular units based on teacher feedback where needed.
  - The grades 6–12 Standards and Curriculum Team developed 24 total curricular units, including one fall and one spring curricular unit in each subject area: grade 6 science, grade 6 math, grade 7 science, grade 7 math, grade 8 science, grade 8 math, biology, chemistry, physics, algebra I, geometry, and algebra II. At the end of each of the first two school years the curriculum was delivered to students, the team met to discuss revisions to the curriculum based on teacher feedback where needed. Included in these revisions was building a "5E" (engage, explore, explain, elaborate, evaluate) exemplary lesson sequence for the first key concept in each STEM curricular unit.
- **Established a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites.** As described in the section describing program components (Section 1.1), CTAC and TUSD established work teams to support the project. Teams were established by the start of the 2019–20 school year, with the exception of those teams focused on implementation that began in Cohorts 2–4. Teams met regularly throughout the year to support implementation.

- The *Leadership Council* comprised six members from CTAC and TUSD who directed and oversaw project implementation. Included in this team were the grant's two co-Project Directors, one each from CTAC and TUSD, as well as TUSD's superintendent and CTAC's CEO. This team met approximately eight times per year.
- The *PreK–5 Standards and Curriculum Team* included three TUSD teachers from each of grades PreK–5. It was established in the 2018–19 school year. The *6–12 Standards and Curriculum Team* was established the following year and included 21 lead teachers spanning grade levels and math and science content areas. Most content teams had two members, except for grade 8 Science, Geometry & Algebra II, and Computer Science, which had one member each. These teams received training throughout their first year of establishment (on, e.g., engineering, computer science, content progression, assessment, project-based learning) and were subsequently responsible for developing and revising the curricular units for their respective grade spans.
- The *PreK–5 District Implementation Team (DIT)*, established in 2019–20, and a separate *6–12 DIT*, established in 2020–21, comprised district administrators and teacher leaders who supported implementation of project-based learning curriculum and pedagogy. These teams met approximately five times per year.
- Each school developed a *Site Implementation Team (SIT)* with at least one administrator and two teachers. These teams supported implementation of project-based learning curriculum and pedagogy and met approximately one–two times per month during the school year, though this frequency varied both by school and by year. The focus of SIT meetings varied across teams in response to school needs; example content included unpacking challenges to program implementation, classroom walkthroughs, planning teacher training, and organizing STEM materials.
- The *Community Collaboratory* represented a diverse set of STEM industry leaders from Tracy's public and private sectors, who assisted with creating field experiences and aligning the curriculum to industry standards. This team met approximately four times per year.
- The *Technical Working Group (TWG)* represented a group of experts from STEM-focused industries and PreK–12 education who discussed implementation, connected students with STEM professionals, and advised the Leadership Council on developments in STEM that could improve program implementation. This team met approximately twice per year.

* **Provided professional support structures and capacity building to effectively implement the curricular units.** CTAC and TUSD provided schools and teachers with various support throughout implementation to best help development, revision, and school use of the curricular units.
  - Schools' SITs received professional development days or pedagogical training. Examples of support included discussing data collection from walkthroughs, co-facilitating school workshops with the SIT, and assisting with planning and lesson studies. School administrators received an additional 3–8 trainings each year.
  - Each year, SITs at participating schools created and implemented a site plan for STEM instruction to support teachers' use of the curricular units. Examples of support implemented within schools included creating a science lab for teachers and students, hosting science nights for families, and implementing rich STEM experiences.
  - District executive leadership worked to engage district or site leaders in bimonthly STEM conversations to message the importance of the initiative and identify support and barriers to school use of the curricular units.

- DITs received approximately five program training sessions throughout each year of program implementation and, in turn, provided at least four training sessions on the curricular units for teachers and school administrators each year.

3.2 Pandemic-related Adaptations

The COVID-19 public health emergency significantly affected implementation of the PreK–12 STEM Pathway for Cohort 1 during the spring of school year 2019–20 and the entire 2020–21 school year. During this time, the curriculum was mostly implemented online by teachers rather than in-person in the classroom and professional learning for teachers in TUSD also was conducted remotely.

During this period of remote learning, the PreK–5 Standards and Curriculum Team made modifications to curricular units for virtual implementation of the program under the guidance of the project co-directors. District leaders and CTAC prioritized materials in each STEM unit that were distributed to students to conduct the units in a distance learning environment. CTAC and TUSD rapidly adjusted the PreK–12 STEM curriculum and trainings for teachers to provide the intervention remotely. For example, each unit includes a hands-on engineering design challenge. CTAC and TUSD adapted the design challenges for students to be able to complete the units in a remote context. This involved making the design challenges safe for home implementation (e.g., in the absence of a teacher's in-classroom supervision) and determining what materials students would need, regardless of their socioeconomic status, to complete the engineering design challenge. TUSD administrators and teachers then worked to ensure all students had the required design challenge materials, even hand-delivering materials to students' homes. Further, TUSD also ensured all students received a device for instruction if they did not already have one. In doing so, CTAC and TUSD sought to guarantee equitable access to the PreK–12 STEM curriculum even during remote learning.

4. Impact Study

The impact study looked at the effects of the PreK–12 STEM Pathway program on English language arts (ELA), math, and science achievement (collectively, "academic achievement") and on two college readiness outcomes (pass rates of Advanced Placement [AP] tests and rates of college preparedness) after two years of exposure to the PreK–12 STEM Pathway. In addition, the impact study looked at the effects on academic achievement after three years of exposure and at effects after two years for select student subgroups (multilingual learners ["MLL students"], Hispanic students, boys, and girls). To examine these outcomes, the impact study relied on publicly available data from the California Department of Education from school years 2014–15 through 2022–23, using a C-SITS design and matching PreK–12 STEM Pathway schools to schools not receiving PreK–12 STEM Pathway during that school year.

Key Impact Study Finding

Overall, the impact study was unable to detect impacts of the PreK–12 STEM Pathway in academic achievement or college readiness. This lack of overall findings was expected, given that the COVID-19 pandemic limited the ability of the study design—put in place prior to implementation—to detect impacts of the intervention.

Subgroup analyses revealed noteworthy variation in findings worthy of attention and further explanation.

4.1 Impact Study Sample

The impact study sample includes 16 TUSD schools serving mainstream PreK–12 students. Of these 16 schools, 12 served as the treatment group. The other four schools were included in the comparison group, along with additional comparison schools from outside of TUSD.[9]

4.1.1 Treatment Schools

TUSD and CTAC selected 12 of the 16 mainstream TUSD schools to implement the PreK–12 STEM Pathway as part of the impact study's treatment group.

- Cohort 1, which included all grades PreK–5 in the treatment schools, began implementation in the 2019–20 school year. Cohort 1 contributed three years of outcome data (school years 2019–20, 2020–21, and 2021–22).
- Grades 6–12 received the PreK–12 STEM Pathway as part of Cohort 2, beginning in the 2021–22 school year. Cohort 2 contributed two years of outcome data (school years 2020–21 and 2021–22).[10]

When selecting treatment schools within TUSD, both CTAC and TUSD staff considered a school's grade configuration, three prior years of ELA and math achievement test scores, socioeconomic disadvantage

[9] TUSD and CTAC preferred to provide the intervention to as many TUSD students as quickly as possible, to allow equitable access to the intervention. Four TUSD schools remained in the comparison condition to allow for the impact study to meet the U.S. Department of Education's What Works Clearinghouse standards with reservations. This level of rigor was required by the grant funding both the implementation and evaluation. These four TUSD comparison schools received the intervention upon the conclusion of the impact study.

[10] Three of the 12 treatment schools served grades PreK–8. Even in these three schools that included both grade bands, grades PreK–5 received the intervention as part of Cohort 1, and grades 6–8 received the intervention as part of Cohort 2.

(proxied by percentage of students receiving Free or Reduced-Price Lunch), and percentage of English learners. Staff were careful to select treatment schools such that the remaining comparison schools did not have outlier or extreme values on any of these dimensions. Exhibit 4-1 shows the treatment and comparison schools used in the study from within TUSD.

Exhibit 4-1. Impact Study Cohorts within Tracy Unified School District

Cohort	Treatment Schools	
	Grades	Years Outcome Data in Impact Study
Cohort 1	PreK–5	2019–20 2020–21 2021–22
Cohort 2	Grades 6–12	2021–22 2022–23
Total	**12 schools**	

4.1.2 Comparison Schools

Two TUSD schools contributed PreK–5 data to the comparison sample for Cohort 1; two TUSD schools contributed grades 6–8 data and one TUSD school contributed grades 9–12 data to the comparison sample for Cohort 2.[11] Additionally, the study team identified non-TUSD comparison schools through a matching process. These schools, together with the four TUSD comparison schools, constituted the comparison group.

In addition to the comparison schools within TUSD, the study team selected, for each treatment school, at least two additional non-TUSD comparison schools from across the state of California to increase statistical power. The study team used baseline demographic and academic achievement data to identify matches and conducted the matching process separately for each cohort and for each impact analysis (as per the research questions in Section 2). In doing so, the study team aimed to achieve a comparison group that was as similar as possible to the treatment group, based on various characteristics of the schools measured prior to the start of the PreK–12 STEM Pathway implementation. Having treatment and comparison groups that are similar on these baseline characteristics increases the likelihood that any differences in outcomes the study team finds can be attributed to the PreK–12 STEM Pathway, rather than to underlying differences between those two groups.

For each achievement outcome, the study team was able to match each PreK–12 STEM Pathway school with two to five comparison schools. For each college readiness outcome, the study team was able to match each treatment school with 10 comparison schools. Comparison condition sample sizes varied between analyses for a few reasons:

- The treatment group to which we matched the comparison group varied between analyses (e.g., college readiness outcomes were only estimated in high schools; three-year impacts were only estimated in elementary schools, etc.).
- TUSD comparison schools sometimes did not match through the systematic procedure. When that happened, the study team hand-matched the TUSD comparison schools to ensure that those schools were part of the comparison group, in addition to the systematically matched comparison schools, creating a comparison group slightly larger than expected.

---

[11] One of the comparison schools serves grades PreK–8, and therefore contributed data to both the Cohort 1 and Cohort 2 comparison samples.

- Achievement data were not systematically available at all California schools in all study years. In the 2020–21 school year, California districts were allowed to opt out of administering the state standardized Smarter Balanced and California Science Test (CAST) tests due to the COVID-19 pandemic. This resulted in high rates of data missingness for academic achievement outcomes in 2020–21 (ETS, 2023a; 2023b). Because comparison schools in both cohorts needed to have non-missing data in 2020–21 (the second year of outcome data for Cohort 1 and baseline year for Cohort 2), this missingness limited the pool of eligible comparison schools for the academic achievement outcome samples.

See Appendix C for more information on the matching process; Appendix D provides further information about the matching results.

4.2 Measures

This impact study used publicly available data downloaded from the California Department of Education website. These data are not publicly available at the individual student level, as this would compromise student privacy. Instead, the California Department of Education provides the public with average data, both for the school overall and for individual grades within each school. These school and grade-within-school averages provide the public with information about individual schools' demographics and outcomes without sacrificing individual student privacy.

The study team drew on publicly available data, and therefore used baseline and outcome data averaged at the school and grade-within-school level for each school in the study sample (e.g., a school's average grade 3 Smarter Balanced ELA scores). The study's sample size is therefore driven by the school-grade-average. The narrative of this report largely omit references to these for parsimony; detail around sample sizes is provided in the technical appendix.

Most baseline data were downloaded in fall 2020; outcome data were downloaded as they became available between spring 2021 and spring 2024.

4.2.1 Outcome Measures

The impact study assessed the effects of the PreK–12 STEM Pathway on the following academic achievement and college readiness outcomes:

- **ELA and math achievement** in grades 3–8 and 11. We used state standardized test data from the Smarter Balanced ELA and math exams in grades 3–8 and 11 (California Department of Education, 2024b).
- **Science achievement** in grades 5, 8, 11, and 12. We used the CAST, which has been in use statewide since spring 2019.[12]
- **AP exam pass rates** in grade 12. For each high school, we defined this rate as the number of grade 12 students who earned a 3 ("qualified") or higher on two or more AP exams prior to graduation (College Board, n.d.), which California reports annually (California Department of Education, 2023b). We divided this number by all grade 12 students enrolled at that school to provide a rate of AP exam passing, which accounts for the varying number of students enrolled in each school.
- **California College Prepared rates** in grade 12**.** The University of California (UC) and the California State University (CSU) systems have established a uniform minimum set of college-preparatory

12 In California, science is tested in grades 5 and 8 and once in grades 10–12. Although high school students can take the CAST when they are in grade 10, 11, or 12, in some districts, such as TUSD, students typically take the CAST in grade 11 or 12 only.

courses required for admission as a freshman, which are referred to as the A–G course requirements (California Department of Education, 2024c; California Department of Education, 2023a).[13] They also list additional criteria to be "prepared" for college (e.g., performance on standardized assessments, earning college credits).[14] This outcome captures the percentage of students who complete the A–G course requirements with a C minus or better, plus meet one of the additional criteria required to earn California's designation of "Prepared" for postsecondary education. For each high school, we defined this rate as the number of grade 12 students who earned the "Prepared" designation, divided by all grade 12 students enrolled at that school.

The study estimated the effects on all outcomes after two years of student exposure to the PreK–12 STEM Pathway. For grades PreK–5, which began prior to grades 6–12, we also measured impacts after the third year of exposure.

Although the study team did not estimate one-year impacts, the impact model (see Section 4.4 and Appendix C) also included data after the first follow-up year when these data were available.

4.2.2 Baseline Measures

In addition to the follow-up data, the study team downloaded baseline data for each outcome. For ELA and math achievement, baseline data were available dating back to the 2014–15 school year. For science achievement, baseline data were available only beginning in the 2018–19 school year, the first year the CAST was administered. For the college readiness outcomes, the study team used baseline data starting in the 2017–18 school year to ensure the outcomes were reported consistently across baseline years (in baseline data prior to 2017–18, the relevant outcomes were not reported in the same manner).

4.2.3 Demographic Data

This study leveraged demographic data both to support matching TUSD treatment schools to comparison schools and to function as a statistical control in the impact model. These data included school-level percentages of White, Black, and Hispanic students, of students identified as socioeconomically disadvantaged, and of English learners.

4.2.4 Missing Data

After pausing administration of state assessments in school year 2019–20, the California Department of Education allowed districts to choose whether to administer the state standardized Smarter Balanced and CAST tests in school year 2020–21 due to the COVID-19 pandemic. Many districts opted not to administer the state tests, resulting in ELA and math test score data being available for only 24% of the students in California and science achievement data being available for only 16% of students in California that year (ETS, 2023a; 2023b). The resulting large rates of data missingness for ELA, math, and science achievement in 2020–21 limited the pool of eligible comparison schools for ELA, math, and science outcome samples, as comparison schools in both cohorts needed to have non-missing data in 2020–21 (the two-year follow-up time point for Cohort 1 and baseline year for Cohort 2).

---

13 A–G course requirements cover courses in History/Social Studies, English, Math, Science, Languages other than English, Visual and Performing Arts, and College Preparatory Electives (California Department of Education, 2024c).

14 These additional criteria include earning a score of at least a Level 3 on the ELA or math Smarter Balanced Summative Assessment and at least a Level 2 on the other assessment; completing one semester (or the equivalent) of college credit courses with at least a grade of C minus and receive college credits for each course; earning a score of 3 on one AP exam or score of 4 on one International Baccalaureate exam; or completing the Career and Technical Education Pathway (California Department of Education, 2023a).

All analyses include only schools without any missing outcome data in the year immediately prior to implementation or in the follow-up year. When a school was missing any key outcome or baseline data needed for the analysis, the study team excluded the school from the analysis entirely. It did not impute any outcome or baseline data. Appendix C.3 provides more information about the number of schools excluded from the analysis due to missing data or other reasons (for example, because the school was a charter school or a special education, juvenile justice, or other type of alternative school).

4.2.5 Outcome Summary

Exhibit 4-2 summarizes the outcomes the study examined and their domains, years of data collected, follow-up period for the impact analysis, and populations for which impacts were estimated.

Exhibit 4-2. Study Domains, Outcomes, Data Years, Follow-up Periods, and Populations

Outcome Domain[a]	Measure	Years of Data	Follow-up Period for Impacts	Population
Literacy Achievement	Smarter Balanced ELA score	2014–15 to 2022–23	Two years	All students (Cohorts 1 & 2)
				MLL students
				Hispanic students
				Boys
				Girls
			Three years	All elementary students (Cohort 1 only)
Math Achievement	Smarter Balanced math score	2014–15 to 2022–23	Two years	All students (Cohorts 1 & 2)
				MLL students
				Hispanic students
				Boys
				Girls
			Three years	All elementary students (Cohort 1 only)
Science Achievement	California Science Test (CAST)	2018–19 to 2022–23	Two years	All students (Cohorts 1 & 2)
				MLL students
				Hispanic students
				Boys
				Girls
			Three years	All elementary students (Cohort 1 only)
College Readiness	Advanced Placement (AP) exam pass rate	2017–18 to 2022–23	Two years	All grade 12 students (Cohort 2 high schools only)
	California College Prepared rate		Two years	All grade 12 students (Cohort 2 high schools only)

[a] Outcome domain definitions are based on the What Works Clearinghouse Study Review Protocol, Version 5.0 (U.S. Department of Education,2023).

4.3 Comparison Condition

In estimating the impacts of any intervention, we compare outcomes for the treatment group to the outcomes of a comparison group. Ideally, we have created treatment and comparison groups that are similar prior to the intervention, and during the intervention period, the only systematic difference between the two groups is the treatment condition itself. Understanding the differences in what the treatment and comparison conditions experienced during the intervention can therefore help us understand to what the estimated treatment effect can be attributed.

During the study period, comparison schools did not have access to the PreK–12 STEM Pathway program.[15] Instead, during the study period, comparison schools offered science, math, or STEM instruction, professional development, and support without access to the PreK–12 STEM Pathway program. Statewide, this comparison condition would include support for the Next Generation Science Standards (NGSS), which the California Department of Education adopted and created support for in 2013 with the goal of improving science teaching and learning across the state. Within TUSD, comparison schools continued to implement content area instruction based on the district-level work that was done through the NGSS Early Implementers Initiative and to receive support from the district to do that work. During the study period, TUSD comparison schools did not implement the STEM interdisciplinary curricular units developed through the PreK–12 STEM Pathway project, nor did they receive the related support that the treatment schools received. Following the study period, comparison schools within TUSD received delayed treatment. Grades PreK–5 in comparison schools within TUSD were provided with the PreK–12 STEM Pathway program in the 2022–23 school year after all study outcome data had been collected for Cohort 1. Similarly, grades 6–12 in comparison schools within TUSD received the PreK–12 STEM Pathway program in the 2023–24 school year after all data had been collected for Cohort 2.

Because the study team selected additional comparison schools outside of TUSD using publicly available data, it was not able to track information on specific services received by those other comparison schools. However, given the increasing focus on STEM in education, the business-as-usual services available to those other schools outside of TUSD also could have included STEM-focused programming. For example, the afterschool STEM program Girls Who Code (Thomas, 2023) is offered to both treatment and comparison high schools in TUSD and other districts. The California STEM Network (n.d.) provides a directory of other STEM programs for students.

Additionally, COVID-19 affected all schools and districts, across both treatment and comparison groups. Further, schools and districts across California responded in varied ways. This included decisions regarding school closures and transitions to virtual learning, as well as changes in instruction and support provided to students and families (Hurtt et al., 2021). Different policies, procedures, and support have the potential to affect student outcomes. Hence, it may be difficult to disentangle potential effects of the PreK–12 STEM Pathway from the ways in which the public health emergency might have affected TUSD treatment schools compared to TUSD and non-TUSD schools in the comparison group.

4.4 Analytic Approach

In research, randomized controlled trial designs typically represent the gold standard of design. These designs designate at random certain study members (e.g., schools or individuals) to receive a program (treatment group), whereas the remaining study members do not receive the program (control group). Because receiving the program or not is random, these designs have the least risk of bias based on unobserved differences between the treatment and control groups.

Because random assignment was not possible for this study, this impact study used a quasi-experimental design to look at the effects of the PreK–12 STEM Pathway. Specifically, the study team used a C-SITS design with a baseline linear trend projection model. In essence, for each outcome, this design takes

[15] Typically, this would be referred to as a "business-as-usual" comparison condition. Given the COVID-19 context of this study, this typical term may feel less apt, but the broader concept holds.

advantage of trends in the outcome for the treatment and comparison groups in the period leading up to the treatment (the baseline years) and uses this information to help estimate impacts in follow-up years.[16]

A C-SITS baseline linear trend projection model uses multiple years of baseline data for both the treatment group and the comparison group to predict a trend for each group. Then the model examines how the actual outcome for each group at follow-up differs from the outcome value that was predicted for that group, based on that group's baseline trend. In other words, the model considers (1) how much the treatment group's actual outcome differed from the value predicted based on the treatment group baseline trend and (2) how much the comparison group's actual outcome differed from the value predicted by the comparison group baseline trend. Finally, the model estimates the difference between (1) and (2); this difference is the estimated impact. The underlying assumption is that, if the treatment group had not started receiving the treatment, the treatment group's baseline trend would have continued into the post-treatment period. Exhibit 4-3 provides a simple example of how the model works, using fictional data.

Exhibit 4-3. Illustration of Comparative Short Interrupted Time Series Baseline Trend Projection Model

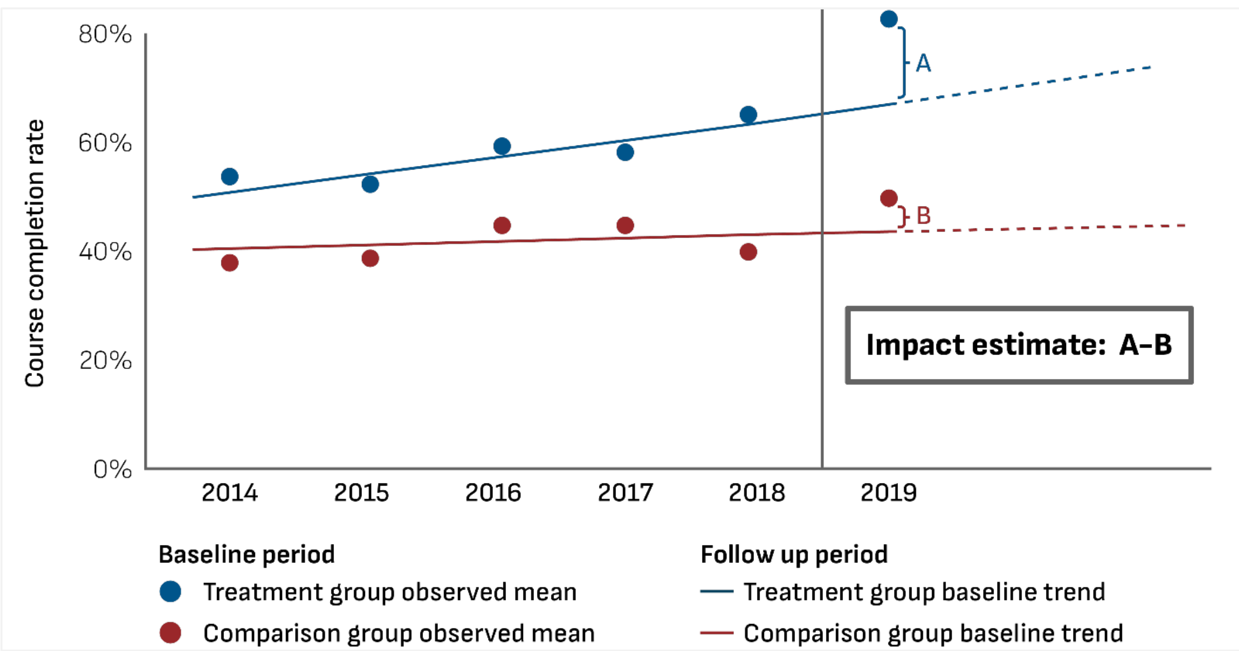


Exhibit adapted from Figure 4.2 in Somers et al. (2013).

Previous studies (Somers et al., 2013; St. Clair et al., 2014) have shown that C-SITS approaches can nearly replicate the results from more rigorous designs, including randomized control trial studies. We do note that one underlying assumption of the C-SITS baseline linear projection model is that, in the absence of the program of interest, baseline trends would have otherwise continued (Hallberg et al., 2018). However, that assumption could be challenging to justify during a pandemic that caused widespread disruption in classrooms across the country (Feller & Stuart, 2021).

---

[16] Baseline linear trend model versions of C-SITS require at least four years of baseline data were available. For science achievement, only one or two years of baseline data were available for each cohort. For these estimates, the study team used a slightly modified model, using baseline mean projection instead of linear trend projection. The two models are similar except the baseline mean projection model assumes that differences between the treatment group and comparison group do not change over time during the baseline years, similar to a difference-in-differences model (Hallberg et al., 2018).

C-SITS designs are often used when multiple years of baseline data are available, as this will allow the model to predict a baseline trend for each group. These multiple baseline years give the model greater statistical power to detect effects. The impact model also includes terms for matching blocks (indicators for each treatment school and its corresponding matched comparison schools) and controls for school-level demographics. Appendix C contains the full impact model.

The academic outcome data for grades 3–8 and 11 for ELA and math and grades 5, 8, and 10, 11, or 12 for science in the current study were available as school's average scores within grade level and year (e.g., a school's average grade 3 ELA Smarter Balanced score in spring 2021 would have been a single data point). To allow for comparisons across grades and years, the study team standardized these school-by-grade-by-year observations using the student-level statewide score means and standard deviations for the relevant grade and testing year. We then analyzed data across grades and years to calculate a single, combined impact estimate across all relevant grades (Somers et al., 2013). Because the achievement outcomes were standardized prior to analysis, the estimated treatment effects can be interpreted as an effect size.

Both college readiness outcomes—AP pass rates and California College Prepared rates—are measured in a single grade. As a result, we did not standardize these outcomes.

Although the impact model factors in multiple years of baseline data and controls for school-level demographics, the impact study cannot completely eliminate the possibility that the treatment and comparison groups might differ in other ways than whether they received the PreK–12 STEM Pathway. As a result, due to the quasi-experimental nature of the design, it is possible that factors other than the PreK–12 STEM Pathway could account for any differences between the outcomes of the treatment and comparison groups. Notably, Cohort 1 includes nine of the 12 treatment schools in the sample, and its first two years of exposure to the PreK–12 STEM Pathway—school years 2019–20 and 2020–21—were the most impacted by COVID-19. The treatment estimate will therefore reflect differences between Cohort 1 schools and comparison schools due to both exposure to the treatment condition and to differences in how schools, districts, and communities were impacted by the pandemic during these years.

4.5 Baseline Equivalence

For the impact study to produce credible evidence of the effects of the PreK–12 STEM Pathway, the treatment and comparison groups should be as similar as possible prior to the start of the PreK–12 STEM Pathway. The study team assessed the extent of this similarity (*baseline equivalence*) for each outcome. For each outcome, the study team assessed baseline equivalence at the school level in the year immediately prior to the start of implementation, which represents the average performance of an earlier cohort of students in the same grades and schools on the same measure. Appendix C provides the baseline equivalence model. The study team achieved baseline equivalence across all samples, with baseline outcome differences of less than 0.05 standard deviations for each outcome. See Appendix D for the baseline equivalence results.

The study team also assessed baseline equivalence on school-level demographics such as race/ethnicity, percentage of students who are socioeconomically disadvantaged, percentage of students who are English learners, and percentage of girls. The results of the school-level demographic baseline equivalence for the confirmatory outcomes are shown in Appendix D. As per the What Works Clearinghouse standards (U.S. Department of Education, 2022), the study team controlled in its impact model for any demographic variables with baseline differences that exceeded 0.05 standard deviations.

4.6 Consideration of Multiple Comparisons

In this study, we provide estimates of 20 different treatment effects across multiple samples, subsamples, time points, and outcome measures (see Exhibit 4-3, above, for a summary), each of which redraws a new

comparison sample. We provide this number of estimated effects to better explore and understand the impacts of PreK–12 STEM Pathway on student impacts. Running this number of outcome estimates comes with an elevated risk of finding results that are statistically significant by chance. Given the number of tests run, we would expect approximately one impact estimate to be statistically significant by chance (Westfall, Tobias, & Wolfinger, 2011; Chen, Feng, & Yi, 2017).

Traditionally, statisticians used a formal multiple comparison correction to deal with this elevated risk of finding a "false positive"/ "false negative" among multiple impact estimates. For example, a Bonferroni correction directs researchers to divide the typical alpha level (the level of the *p*-value against which statistical significance should be assessed) of 0.05 by the number of impacts estimated in the study. In this study, that would suggest that any individual impact estimate should be compared not against the *p*-value of 0.05, but rather against a *p*-value of 0.0025 (calculated as 0.05/20).

This rather simplistic approach to multiple comparison corrections, however, has several downsides, including being overly conservative (that is, unduly elevating the risk of not finding a statistically significant result when a result exists) and ignoring the differences between the multiple tests performed (in terms of, e.g., the relationship between individual analyses, the relative strength of evidence of the theory being tested; Bender & Lange, 2001). More modern extensions of this approach adjust *p*-values differentially, instead of applying the same adjustment to all *p*-values. For example, if we were to use the Benjamini-Hochberg approach, we would rank order the estimated results by *p*-value. In this adjustment, only the smallest observed *p*-value would be assessed against 0.0025 to determine statistical significance. The second smallest observed *p*-value would be assessed against 0.0026 (calculated as 0.05/19), the third smallest against 0.0028 (calculated as 0.05/18), and so on until the largest observed *p*-value, which would be assessed against 0.05 (Chen, Feng, & Yi, 2017). This procedure, similar to the original Bonferroni adjustment, does not take into consideration the relative import of each hypothesis test. Using these procedures to adjust for multiple comparison corrections, therefore, risks limiting researchers' willingness to explore additional research questions. Of particular concern might be an unwillingness to estimate subgroups analyses, which may provide useful insight into the variation of program efficacy.

Given the import of this methodological consideration, the U.S. Department of Education commissioned a panel of 13 leading methodologists in the field of impact evaluations to author guidelines on how researchers should consider the issue of multiple comparison corrections in education evaluations (Schochet, 2008). These recommendations guide our approach to the issue of multiple comparisons in this study.

Specifically, the guidelines recommend identifying confirmatory research questions and associated domains (e.g., impacts on a specific outcome, or for a specific subgroup) prior to the analysis of any data. Researchers should then use multiple comparison corrections when analyzing data for these confirmatory research questions (e.g., performing a meta-analytic procedure to combine multiple impact estimates into a single hypothesis test within a domain). Following these guidelines prior to collection of outcome data, we specified three confirmatory outcome questions related to the impacts of the intervention on all students in PreK–12 STEM Pathway schools after two years of exposure to the program on ELA, math, and science achievement. We specified that, following WWC standards, we considered each of these outcome measures as a separate domain. We therefore do not apply a formal multiple comparison correction in this study.

The guidelines on multiple comparison correction do not recommend using multiple comparison corrections on exploratory research questions but rather provides the following guidance for the interpretation of these results, which were not prespecified as confirmatory in advance:

> *"Results from post hoc analyses are not automatically invalid, but, irrespective of plausibility or statistical significance, they should be regarded as preliminary and unreliable unless they can be rigorously tested and replicated in future studies (Schochet, 2008, p. 4)."*

We, therefore, following these guidelines, do not perform a formal multiple comparison correction on the 17 exploratory research questions in this study. Rather, we interpret the results of these exploratory research questions with a greater level of caution than we do the three confirmatory estimates, considering any statistically significant results to be "preliminary" and worthy of further exploration.

4.7 Program Effects

The study team estimated the effects of the PreK–12 STEM Pathway on academic achievement and college readiness across different populations and time points. This section presents the results of those impact analyses.

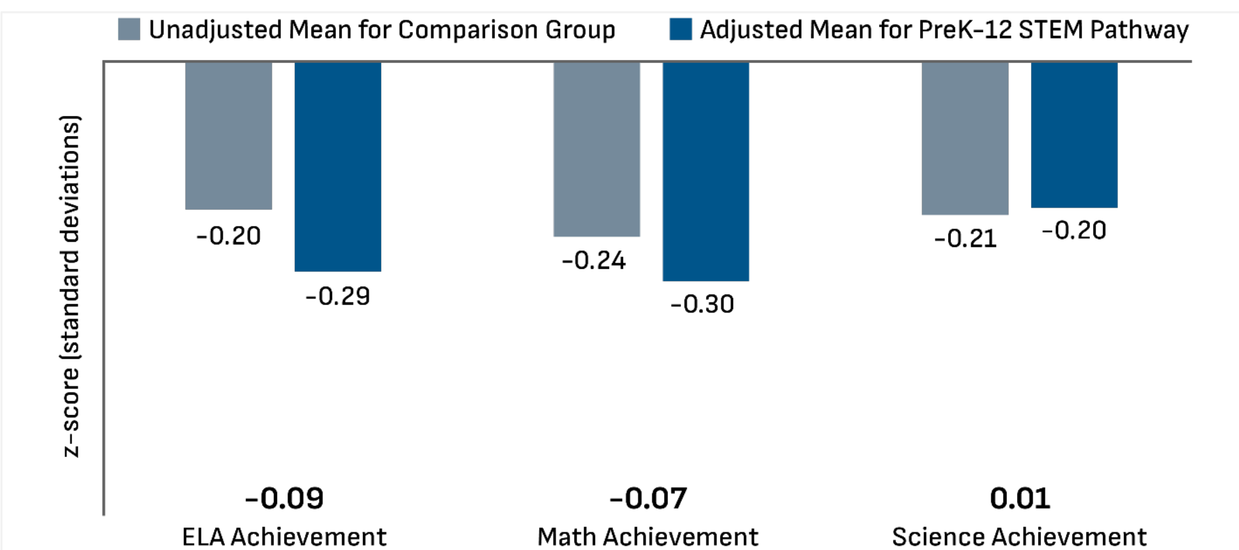
4.7.1 Confirmatory Program Effects: Academic Achievement for All Students after Two Years of Exposure

Exhibit 4-4 presents the effects of the PreK–12 STEM Pathway on academic achievement for all students in Cohorts 1 and 2, after two years of exposure to the PreK–12 STEM Pathway. It presents the mean levels of the outcomes at the follow-up time point (two or three years, depending on the analysis) for comparison schools, as well as the model-adjusted mean for schools that received the PreK–12 STEM Pathway (treatment schools). Academic achievement outcomes are standardized relative to state means.

The PreK–12 STEM Pathway did not have any effects on academic achievement after two years of exposure, when looking at the effects across all students and both Cohorts 1 and 2 (Exhibit 4-4). Although the impact estimates range from -0.01 to -0.09, depending on the outcome, these results are statistically not distinguishable from zero.

Among all the outcomes, the science achievement outcome is perhaps most aligned with the theory of change of the PreK–12 STEM Pathway. However, because the CAST was only offered starting in school year 2018–19, fewer years of baseline data were available. Given that the C-SITS design derives much of its power from multiple years of baseline data, the availability of relatively few years of baseline data for science achievement might have limited the study's ability to detect effects.

Exhibit 4-4. Effects of the PreK–12 STEM Pathway on Academic Achievement after Two Years of Exposure (All Students)

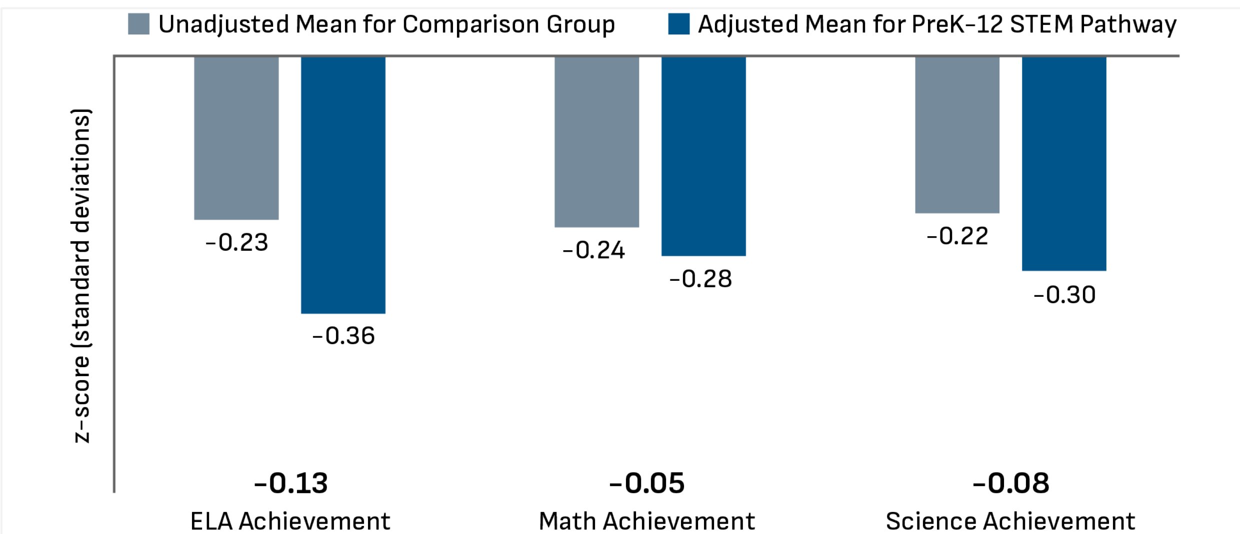


Notes: The comparison group bars show the unadjusted comparison group means for ELA achievement, math achievement, and science achievement, respectively. The PreK–12 STEM Pathway bars show the model-adjusted treatment mean for ELA achievement, math achievement, and science achievement, respectively. Comparison and the PreK–12 STEM Pathway means reflect data for the two-year follow-up point for the respective cohort (2020–21 for Cohort 1 and 2022–23 for Cohort 2). Below the set of bars for each outcome is the impact estimate from the model. The impact estimates might differ slightly from the differences between the displayed intervention and comparison means due to rounding. ELA and math achievement reflect Smarter Balanced ELA and math scores for grades 3–8 and 11 in standard deviation units (z-scores). Science achievement reflects California Science Test scores for grades 5, 8, and one of grades 10–12 in standard deviation units. Data come from the California Department of Education.

4.7.2 Exploratory Effects on Academic Achievement after Three Years of Exposure for All Students in Cohort 1

The study team also did not detect any effects on academic achievement after three years of exposure (Exhibit 4-5). As with effects after two years, the three-year impacts are statistically indistinguishable from zero. Academic achievement after three years could be measured only for Cohort 1, given the study time period and the timing of outcome data release.

Exhibit 4-5. Effects of the PreK–12 STEM Pathway on Academic Achievement after Three Years of Exposure (Cohort 1)



Notes: The comparison group bars show the unadjusted comparison group means for ELA achievement, math achievement, and science, achievement, respectively. The PreK–12 STEM Pathway bars show the model-adjusted treatment means for ELA achievement, math achievement, and science achievement, respectively. Comparison and the PreK–12 STEM Pathway means reflect data for the three-year follow-up point (2021–22) for Cohort 1. Below the set of bars for each outcome is the impact estimate from the model. The impact estimates might differ slightly from the differences between the displayed PreK–12 STEM Pathway and comparison means due to rounding. ELA and math achievement reflect Smarter Balanced ELA and math scores for grades 3–8 and 11 in standard deviation units (z-scores). Science achievement reflects California Science Test scores for grades 5, 8, and one of grades 10–12 in standard deviation units. Data come from the California Department of Education.

4.7.3 Exploratory Effects on Academic Achievement after Two Years of Exposure for Student Subgroups in Cohorts 1 and 2

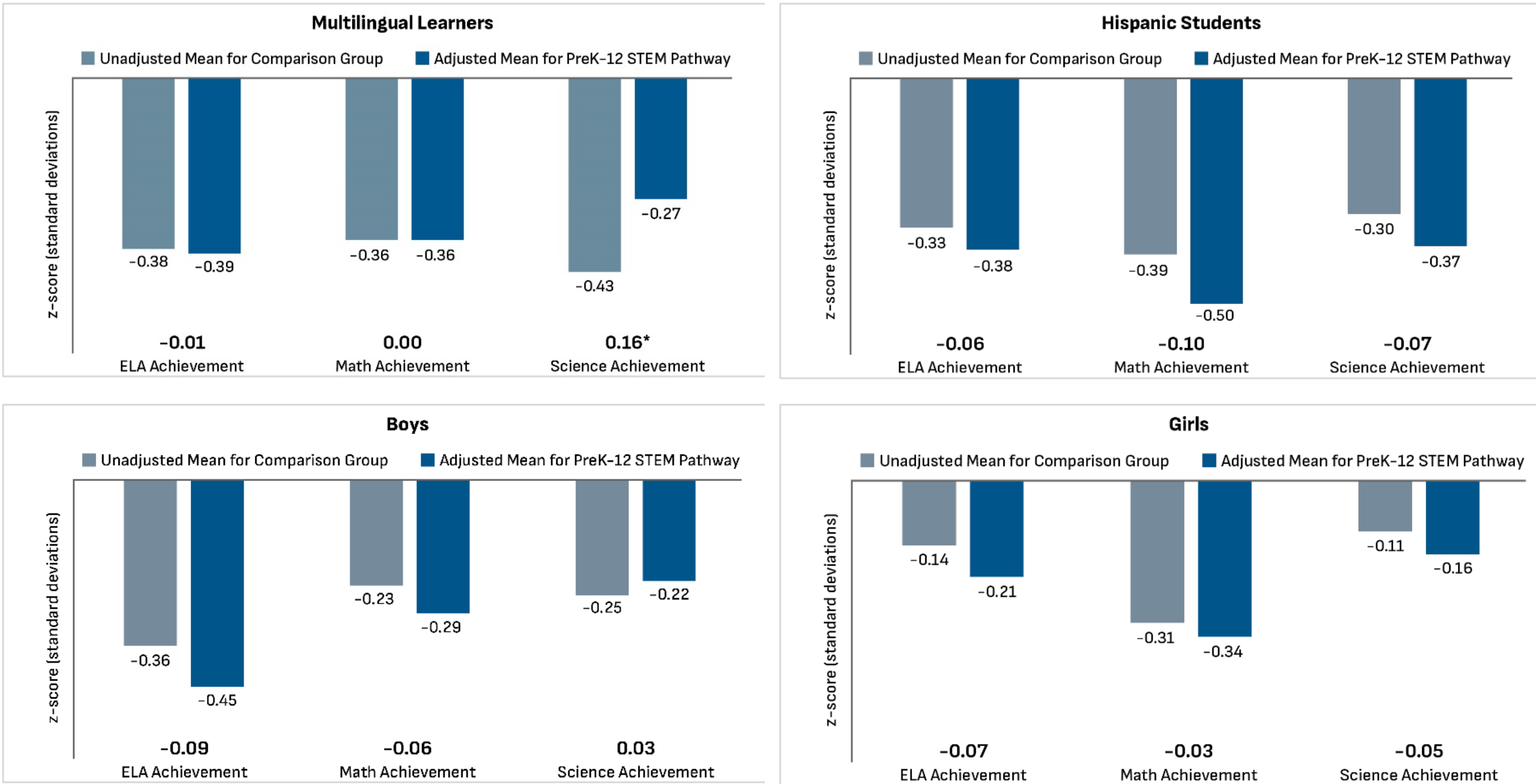
The study also examined the effects of the PreK–12 STEM Pathway on academic achievement after two years of exposure across Cohort 1 and 2 separately for four subgroups: MLL students, Hispanic students, boys, and girls. Exhibit 4-6 presents these results. As with the overall results, the study did not detect any statistically significant effects for any subgroup for ELA or math achievement. However, the PreK–12 STEM Pathway did have a statistically significant, positive effect on science achievement for MLL students ($p < 0.05$). No significant effects on science achievement were detected for other subgroups or for MLL students in ELA or math.

As discussed above (see section 4.6), we interpret these exploratory results as "preliminary," given the high probability of estimating one statistically significant impact among the 20 estimates in this study. We provide some consideration of the rationale behind why the PreK–STEM pathway program may have an impact on MLL students in science, with the aim of developing theory for future studies of the program.

The estimated positive effect on science achievement for MLL students might reflect the program's attempts to be responsive to the needs of subpopulations in TUSD such as MLL students, who often experience disadvantage and who represent a large share of the student population. In addition, the preliminary effects might reflect the additional opportunities for learning in STEM content areas—for example, through the PreK–12 STEM Pathway's STEM curricular units—offered by the PreK–12 STEM Pathway. Moreover, having students complete activities, rather than rely on language, to build mental models—as per the "Activity Before Concept, Concept Before Vocabulary" model (Konicek-Moran & Keely, 2015)—might be an especially helpful learning strategy for English learners whose language skills are still developing. As such, the PreK–12 STEM Pathway's project-based learning opportunities, as well as emphasis on real-world applications of engineering and other topics, might have been especially helpful for providing hands-on and engaging activities to English learners.

In addition to the factors noted above, the PreK–12 STEM Pathway's project-based learning opportunities and other curricula might have made a difference particularly during COVID-19, when other districts might have struggled to meet the needs of MLL students. Evidence suggests that English learners might have been particularly adversely affected by the pandemic. As schools shifted to remote learning, English learners often lacked support and resources to foster their learning and English language development, as well as to address other specific needs (Capers & Morales, 2024; Esquivel, 2021). For example, during the pandemic, many English learners had few opportunities to develop their language skills by speaking in English to their classmates and teachers; moreover, scaffolding techniques often used in in-person settings to support learning, such as small group discussion, might have been harder to replicate remotely (Capes & Morales, 2024).

Exhibit 4-6. Effects of the PreK–12 STEM Pathway on Academic Achievement after Two Years of Exposure, Selected Subgroups



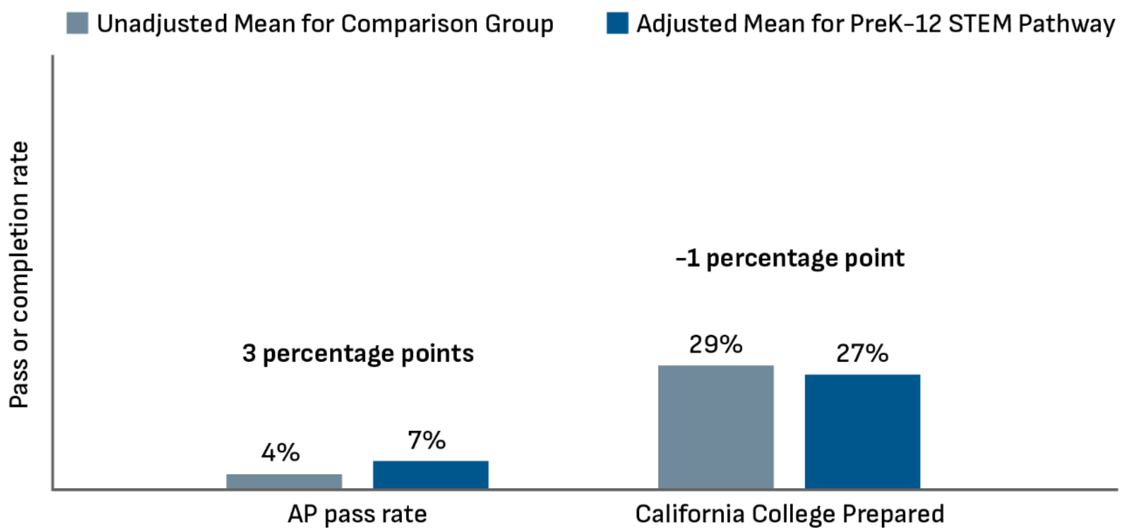
Notes: In each graph, the comparison group bars show the unadjusted comparison group means for ELA, math, and science achievement, respectively, and the PreK–12 STEM Pathway bars show the model-adjusted treatment means for ELA, math, and science achievement. In each graph, comparison and PreK12–STEM Pathway means reflect data for the two-year follow-up point for the respective cohort (2020–21 for Cohort 1 and 2022–23 for Cohort 2). Below each set of bars for each outcome are the impact estimates from the model. Each impact estimate might differ slightly from the difference between the displayed PreK–12 STEM Pathway and comparison means due to rounding.
One asterisk (*) signifies the result is statistically significant at the $p < 0.05$ level. ELA and math achievement reflect Smarter Balanced ELA and math scores for grades 3–8 and 11 in standard deviation units (z-scores). Science achievement reflects California Science Test scores for grades 5, 8, and one of grades 10–12 in standard deviation units. Data come from the California Department of Education.

4.7.4 Exploratory Effects on College Readiness Outcomes After Two Years of Exposure

In addition, the study looked at the effects of the PreK–12 STEM Pathway on AP pass rates and California College Readiness rates after two years of exposure. As shown in Exhibit 4-7, the impact study did not detect any significant effects on either outcome.

The college readiness outcomes were assessed only for schools serving grade 12 in Cohort 2, given the nature of the outcomes. Only two treatment schools served grade 12. Due to the small number of treatment schools that needed to be matched for these outcomes, for each of these outcomes, the study team was able to find 10 comparison schools to match to each treatment school (see Appendix D). Still, relative to the other outcomes, the small number of treatment schools available limited the study team's ability to detect effects for the two college readiness outcomes.

Exhibit 4-7. Effects of the PreK–12 STEM Pathway on College Readiness after Two Years of Exposure



Notes: The comparison group bars show the unadjusted comparison group mean for California College Prepared rate and AP pass rate, respectively. The PreK–12 STEM Pathway bar shows the model-adjusted treatment mean for the two treatment schools across both outcomes. The comparison and the PreK–12 STEM Pathway means reflect data for the two-year follow-up point (2022–23) for Cohort 2 schools. Above the set of bars for each outcome is the impact estimate from the model. In each impact estimate. The impact estimate might differ slightly from the difference between the displayed PreK–12 STEM Pathway and comparison means due to rounding. The California College Prepared rate reflects the percentage of students who earned "Prepared" by completing A–G course requirements and an additional requirement; the AP pass rate reflects the percentage of students who earned a passing score (3 or higher) on at least two AP tests. Data come from the California Department of Education.

5. Conclusion

CTAC and TUSD developed and implemented the PreK–12 STEM Pathway in TUSD at a time when STEM skills are increasingly in demand in the workforce, yet elementary, middle, and secondary student mathematics scores have been in decline (National Science Board, 2024) and women and Hispanic/Latinx, Black, American Indian, and Alaska Native people have been underrepresented in STEM occupations (National Science Board, 2024). CTAC and TUSD undertook this work with the aim of providing participating students with equitable access to rigorous, relevant, and engaging STEM experiences.

5.1 Summary

To implement this program, TUSD and CTAC developed a multi-tiered system of leadership teams, including instructional, administrative, and teacher leaders from the district, each school, and every grade and STEM content area to develop, revise, and train the districts' teachers on 52 multidisciplinary, hands-on, project-based curricular units for each grade and STEM content area in the district. Teachers received ongoing support from the Site Implementation Teams (SITs), site administrators, and District Implementation Teams (DITs) to facilitate implementation. This network of leadership teams was supplemented and supported by district administration, CTAC, local industry leaders, and national experts in the field. A key aspect of the PreK–12 STEM Pathway was placing the STEM curricular units directly into the core curriculum, providing teachers with integrated learning materials to teach hands-on, project-based approaches to computer science and engineering.

The PreK–12 STEM pathway was initially rolled out to grades PreK–5 in the 2019–20 school year and was therefore temporarily disrupted during the TUSD transition to online instruction; online instruction continued throughout spring of 2020 and the entire 2020–21 school year. The district's multitiered system of leadership teams facilitated the flexibility of the intervention to adopt to a distance learning context. Teams modified the Pre-K–12 STEM Pathway's grades PreK–5 integrated curricular units for virtual delivery of the program. Included in these adaptations, teams worked with district leaders to determine and distribute materials to students at home to provide students with hands-on experiences, despite the virtual learning environment. and worked to distribute materials to students at home. Concurrent with the resumption of in-person learning, the leadership team rolled out the PreK–12 STEM Pathway to grades 6–12 in the 2021–22 school year.

5.2 Limitations and Considerations

As part of this study, we conducted an impact study to estimate the effects of the PreK–12 STEM Pathway on student outcomes. Our confirmatory research questions asked about the impacts of PreK–12 STEM Pathway on students' ELA, math, and science achievement after two years of program exposure: our statistical models did not estimate an impact on any of these outcomes. We also estimated 17 additional exploratory research questions, asking about the impacts of the PreK–12 STEM Pathway on academic achievement after three years, on college readiness after two years, and on academic achievement for four subgroups after two years.

Of these 17 estimated exploratory effects, one impact estimate—the estimated effect of the PreK–12 STEM Pathway on MLL students' science achievement—was positive and statistically significant. Consistent with standard practice for interpreting one significant result among 20 total comparisons, we take this finding as preliminary and worthy of further research. Future work that looks to explore these preliminary impact estimates of the PreK–12 STEM Pathway on MLL students in science achievement may consider looking to the integrated and project-based nature of the curricular units developed within this grant. In particular, having students complete activities, rather than rely on language, to build mental

models might be an especially helpful learning strategy for English learners whose language skills are still developing (see, e.g., Konicek-Moran & Keely, 2015).

As with any research, this study includes limitations. The onset of the disruption caused by the COVID-19 public health emergency coincided with the initial two years of Cohort 1's intervention period and had a large impact on program implementation and the data available for analysis. Although COVID-19 affected all schools and districts, across both treatment and comparison groups, schools and districts across California responded in varied ways. This included decisions regarding school closures and transitions to virtual learning, as well as changes in instruction and support provided to students and families (Hurtt et al., 2021). Different policies, procedures, and support have the potential to affect student outcomes. Hence, it is difficult to disentangle potential effects of the PreK–12 STEM Pathway from the ways in which the public health emergency might have affected TUSD schools compared to schools in the comparison group. Moreover, one underlying assumption of the research design used—and put in place at the study's onset—is that, in the absence of the program of interest, baseline trends would have otherwise continued (Hallberg et al., 2018). However, that assumption could be challenging to justify during a pandemic that caused widespread disruption in classrooms across the country (Feller & Stuart, 2021). Notably, because Cohort 1 includes nine of the 12 treatment schools in the sample, any influence of COVID-19 on Cohort 1 is likely to drive the overall program effects. We also note that, due to the small number of schools administering the state achievement tests during the 2020–21 school year, the comparison pool was smaller than anticipated for both cohorts. This led to fewer matched comparison schools than expected, which diminished statistical power, making it harder to detect effects.

In addition, the study might not have detected any effects given limitations to the outcome measures available in the publicly available data. Specifically, the study team relied on state tests, AP exams, and course completion data that were available on the California Department of Education's website. These broad and more distal outcomes would be less sensitive to program impacts than would a more closely aligned outcome measure. More research that uses other measures that are more sensitive and that align better with the different components of the program could yield valuable evidence of the program's potential effects on these outcomes and other potential outcomes of interest.

Another consideration is that the study measured the effects of the PreK–12 STEM Pathway after two years (for most analyses) or three years (for a smaller number of analyses). In general, districts, schools, teachers, and other staff often face learning curves or encounter challenges when they begin to implement new programs. For some programs, even if implementation takes place as intended, it can take time for the programs to lead to effects in measured outcomes. These factors can make it difficult to detect effects on some outcomes after shorter periods of exposure to the program. It is possible that, had schools been exposed to the PreK–12 STEM Pathway for longer periods, the study might have detected significant effects.

5.3 Looking Ahead

Despite the challenges that arose with the COVID-19 public health emergency, CTAC and TUSD were able to form a strong partnership and leverage district leaders to encourage school and teacher buy-in and to emphasize the importance of PreK–12 STEM Pathway. The PreK–12 STEM Pathway's focus on providing every student within TUSD access to rigorous STEM coursework was reinforced through various levels of support, including district leader support of principals, principal and teacher leader support of other teachers, and teacher support of students and families. Additionally, CTAC's approach of phasing in the development of the curricular units over time helped ensure teams had capacity and time to devote to this ambitious initiative. Momentum from earlier successes as well as district leadership support also helped CTAC and school staff to nimbly adapt and overcome challenges as they arose during implementation. For example, during the COVID-19 public health emergency, the multi-tiered system of leadership teams facilitated adaptation to a virtual learning environment.

At the same time, CTAC and TUSD's experiences also shed light on program needs and potential challenges related to implementing the PreK–12 STEM Pathway. From reflections by the CTAC and TUSD team leadership at the grant's conclusion, these include:

- the importance of districts having systems for sourcing, distributing, and communicating STEM toolkits and other materials to principals and teachers.
- the complexities of developing new curricular units and integrating them into existing programming.
- the importance of shifting the balance of the curriculum's content focus to increase emphasis on computer science and engineering.
- the lower levels of training that teachers in younger grades typically receive in computer science and engineering (as opposed to ELA), such that teachers may need more intensive professional development to feel ready to teach STEM.
- the need to have a system to effectively track and manage data sourced from districts in order to monitor program implementation fidelity and meet other grant requirements.

Lessons learned from this study will inform CTAC's continued efforts to develop and implement high-quality, hands-on STEM experiences for students, as well as have implications for STEM curriculum development and STEM research more broadly.

Technical Appendix

Preface and Overview

We include this technical appendix to the report narrative both to provide additional technical detail and to fulfill evaluation reporting requirements for the grant funding this project. In writing the report narrative we found, at times, that clarity for a broader audience conflicted with some of the more detailed requirements for grant reporting. Where we found this difficulty, we typically erred on the side of clarity and streamlined or reorganized detail in the report narrative. As a result, readers looking to closely compare the report narrative to this technical appendix may find some elements less closely aligned than would be ideal for comparing across the two. To help guide this reader, we here provide an outline of the technical appendix, with some elaboration on what may provide harder to compare across the two documents.

Appendices A and B provide the grant-aligned reporting on implementation fidelity, which includes the most distinct differences from the main text.

- EIR evaluation reports are expected to measure implementation "fidelity." Namely, for each year of implementation and each key component specified in the logic model, evaluators are expected to provide a binary descriptor of whether the program was implemented to a level that would indicate fidelity to the intended model. This analytic approach can be useful for providing a direct and parsimonious way of describing the overall findings. The detail on what is required for implementation with fidelity also provides valuable information to future implementors looking to replicate findings in a report. However, while the top line results may provide useful clarity, the level of detail and data required to fully describe the data and methods can be overwhelming and difficult to navigate. Additionally, given the highly unusual circumstances of implementation during the COVID-19 pandemic, we anticipate the specifics of replication may not be directly transferable to future implementations. We therefore opted to describe the high-level approach to and execution of the implementation without reference to implementation fidelity. We trust that readers who delve into the details of the implementation fidelity will find that we have well-represented the data and findings.

Appendices C and D provide the supporting detail for the impact analyses. These are intentionally aligned to current reporting practices to meet What Works Clearinghouse Standards (version 5). There are two primary distinctions between these sections and the main narrative.

- First, in the main narrative, we made an attempt to unpack and explain key features of the research design that would allow a reader to understand the results of this study. We make less an effort to explain the rational in this technical appendix, opting instead to streamline text to allow a WWC reviewer to find the detail needed to assess this study against those standards.

Second, this impact study analyzed data points composed of average scores, within a grade, within a school, within a year. That is, a PreK–5 school in a single year would have 3 data points in the ELA achievement analyses—one each for that year's average scores in grades 3, 4, and 5. This data structure was the best decision for analysis, allowing for relatively greater statistical precision. However, it also provides a struggle in discussing sample sizes ("n of grade-within-school-within-year averages"). We opted for parsimony and accessibility in the main text by removing discussion of sample sizes and largely avoiding the discussion of the data structure. We include these sample sizes within the technical appendix, and err on the side of precision, not parsimony, in describing the sample and analyses.

Appendix A. Fidelity of Implementation Matrix

Abt measured fidelity of implementation for each of the PreK–12 STEM Pathway's three key components. Abt, CTAC, and TUSD developed a fidelity matrix specifying thresholds for adequate implementation for each of the three key components of the PreK–12 STEM Pathway intervention (Exhibit A-1). Fidelity data comes from multiple sources, including training and meeting attendance records, school site plans, and curriculum documents. CTAC collected fidelity data at the end of school years 2019–20, 2020–21, 2021–22, and 2022–23 and provided the data to Abt for analysis.

Abt analyzed fidelity of implementation for each key component by rolling up indicator scores at the unit level to the sample (program) level and comparing them to the defined threshold for adequate implementation. Some indicators were measured only in certain years of implementation. Scoring for each key component is described below.

Key Component 1: Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science

Key component 1 had five indicators: (1) Standards and Curriculum Teams are established, (2) Standards and Curriculum Teams are trained, (3) PreK–5 STEM project-based learning curricular units are developed, (4) Grades 6–12 STEM project-based learning curricular units are developed, and (5) Curricular revisions are made.

1. Indicator 1 required the program to establish a Standards and Curriculum Team to create the curricular units for grades PreK–5 and grades 6–12. The program received one point if the team was established and zero points if the team was not established. This indicator was measured in school year 2019–20 for the PreK–5 team and school year 2020–21 for the grades 6–12 team.

2. Indicator 2 required that at least one of the representatives from each grade level on the Standards and Curriculum Teams be present at 80% or more of the Standards and Curriculum Team training sessions. The program received one point if at least one grade-level representative was present at 80% or more of training and zero points if at least one grade-level representative was present at fewer than 80% of training. This indicator was measured in school year 2019–20 for the PreK–5 team and school year 2020–21 for the grades 6–12 team.

3. Indicator 3 required the PreK–5 team to develop at least two STEM curricular units per grade level, with a maximum of four curricular units per grade level. The program received one point if at least 14 curricular units were created, with at least two curricular units per grade level and zero points if 13 or fewer curricular units were created or if at least two curricular units were not created for each grade level. This indicator was measured in the school year 2019–20.

4. Indicator 4 required the grades 6–12 team to develop 12–24 STEM curricular units, with at least five integrated curricular units to be taught in grades 7–12 math courses, five curricular units to be taught in grades 7–12 science courses, and at least two curricular units in grade 6. The program received one point if the team met these criteria and zero points if any of the criteria was not met. This indicator was measured in the school year 2021–22.

5. Indicator 5 required the Standards and Curriculum Teams to meet at least once to discuss revisions to the curricular units during the first two years the units were delivered by teachers. The program received one point if the team met at least once per year to discuss revisions and zero points if the team did not meet to discuss revisions. This indicator was measured in school years 2019–20 and 2020–21 for the PreK–5 team and in school years 2021–22 and 2022–23 for the grades 6–12 team.

The key component level threshold was met if the program received one point for each indicator measured that school year.

Key Component 2: Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites

Key component 2 had six indicators: (1) The program establishes work teams, (2) Community Collaboratory meets on a regular basis, (3) District Implementation Teams (DITs) meet on a regular basis, (4) Site Implementation Teams (SITs) meet on a regular basis, (5) Technical Working Group (TWG) meets on a regular basis, and (6) Leadership Council meets on a regular basis.

1. Indicator 1 required the program to establish five sets of work teams: the Community Collaboratory, consisting of members from the community who assist with creating field experiences; the DITs, consisting of district administrators and teacher leaders who support implementation of the curricular units; the SITs, consisting of a school administrator and teachers who support implementation of project-based learning curricular units and pedagogy at each school; the TWG, consisting of experts from STEM-focused industries and PreK–12 education who advise the Leadership Council on developments in STEM that can improve implementation; and the Leadership Council, consisting of members who direct and oversee project implementation.

   The program received one point for each team established. This indicator was measured in:

   - School year 2019–20 for establishment of the Community Collaboratory, the PreK–5 DIT, the Cohort 1 PreK–5 SITs, the TWG, and the Leadership Council (for a total of five possible points).
   - School year 2020–21 for the establishment of the grades 6–12 DIT (one possible point).
   - School year 2021–22 for the establishment of the Cohort 2 grades 6–12 SITs (one possible point).
   - School year 2022–23 for the establishment of the Cohort 3 PreK–5 SITs (one possible point).

Indicators 2–6 set metrics for how often each group was required to meet throughout the school year:

2. The Community Collaboratory was required to meet four times during the year. The program received one point if the Community Collaboratory met four or more times and zero points if the Community Collaboratory met fewer than four times. This indicator was measured in all four years of program implementation.
3. The DITs were required to meet at least four times during the year. At the unit level, the team received two points if they met more than four times per year, one point if they met four times per year, and zero points if they met fewer than four times per year. At the sample level, the program received one point if 100% of DITs scored a one or two and zero points if fewer than 100% of DITs scored a one or two. This indicator was measured in all four years of program implementation.
4. The SITs were required to meet at least three times per year. At the unit level, each team received two points if they met more than three times per year, one point if they met three times per year, and zero points if they met fewer than three times per year. At the sample level, the program received one point if 75% or more of SITs scored a one or two and zero points if fewer than 75% of SITs scored a one or two. This indicator was measured in all four years of program implementation.
5. The TWG was required to meet twice per year. The program received one point if the TWG met two or more times and zero points if the TWG met fewer than two times. This indicator was measured in all four years of program implementation.

6. The Leadership Council was required to meet six times per year in 2019–20 and 2020–21 and four times per year in 2021–22 and 2022–23. The program received one point if the Leadership Council met or exceeded the criteria and zero points if the Leadership Council did not meet the criteria. This indicator was measured in all four years of program implementation.

The key component level threshold was met if the program received one point for each indicator measured that school year.

Key Component 3: Provide professional support structures and capacity building to effectively implement the curricular units

Key component 3 had six indicators: (1) School site plans, (2) Training for DITs, (3) Support SITs, (4) Site administrator training, (5) District-level training for teachers, and (6) District executive leadership support.

1. Indicator 1 required schools to establish and implement a site plan for STEM instruction. Schools planned to implement actions within at least three areas of STEM implementation. At the unit level, each school received two points if the plan was developed and implemented, one point if the plan was developed but not implemented, and zero points if the plan was neither developed nor implemented. At the sample level, the program received one point if 75% or more of schools scored two and zero points if fewer than 75% of schools scored a two. This indicator was measured in all four years of program implementation.
2. Indicator 2 required that the PreK–5 DIT and 6–12 DIT receive at least five program trainings throughout the year. At the unit level, a DIT received one point if the team received five trainings and zero points if they received fewer than five trainings. At the sample level, the program received one point if 100% of DITs scored a one and zero points if fewer than 100% of DITs scored a one. This indicator was measured in all four years of program implementation.
3. Indicator 3 required that the SITs receive support from leadership teams at least three times throughout the school year. This support could come in many forms, such as professional development days or pedagogical training. At the unit level, each team received one point if they received at least three supports and zero points if they received fewer than three supports. At the sample level, the program received one point if 75% or more of SITs scored a one and zero points if fewer than 75% of SITs scored a one. This indicator was measured in all four years of program implementation.
4. Indicator 4 required that site administrators participate in at least three training sessions throughout the school year. At the unit level, each site administrator received one point if they participated in three or more training sessions and zero points if they participated in fewer than three training sessions. At the sample level, the program received one point if 75% or more of site administrators scored a one and zero points if fewer than 75% of site administrators scored a one. This indicator was measured in all four years of program implementation.
5. Indicator 5 required that each DIT provide at least four training sessions for teachers about curriculum topics. At the unit level, the DIT received one point if they provided at least four training sessions and zero points if they provided fewer than four training sessions. At the sample level, the program received one point if 100% of DITs scored a one and zero points if fewer than 100% of DITs scored a one. This indicator was measured in all four years of program implementation.
6. Indicator 6 required that district executive leadership engage district and/or site leaders in at least six STEM conversations. The program received one point if district executive leadership met this criterion and zero points if district executive leadership did not meet this criterion. This indicator was

not created until the 2020–21 school year, and thus was measured only during the 2020–21, 2021–22, and 2022–23 school years.

The key component level threshold was met if the program received one point for each indicator measured that school year.

Exhibit A-1. Implementation Fidelity Matrix

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at Sample Level
Key Component 1. Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science			
1. Standards and Curriculum work teams established	Program	1 = Team established. 0 = Team not established.	N/A
2. Standards and Curriculum work teams trained	Program	1 = At least one representative from each grade-level team is present at 80% or more of training sessions. 0= At least one representative from each grade-level team is not present at 80% or more of training sessions.	N/A
3. PreK–5 STEM project-based learning curriculum developed	Program	1 = Program develops 14-28 curricular units, with at least 2 curricular units created for each grade level. 0 = Program develops 13 or fewer curricular units and/or does not develop at least 2 curricular units for each grade level.	N/A
4. 6–12 STEM project-based learning curriculum developed	Program	1 = Program develops 12–24 curricular units, with at least 5 integrated curricular units developed to be taught in math courses and 5 integrated curricular units developed to be taught in science courses for grades 7–12, and at least 2 curricular units for grade 6. 0 = Program develops 11 or fewer curricular units and/or does not develop at least 5 integrated curricular units to be taught in math courses and 5 integrated curricular units to be taught in science courses for grades 7–12, and at least 2 curricular units for grade 6.	N/A
5. Curriculum revisions made	Program	1 = Standards and Curriculum team meets at least 1 time per year to discuss revisions during the first two years of curricular unit delivery. 0 = Standards and Curriculum does not meet to discuss revisions during the first two years of curricular unit delivery.	N/A

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at Sample Level
Key Component 1 Total Score		School year 2019–20: Sum of indicators 1 (Cohort 1, PreK–5), 2 (Cohort 1, PreK–5), 3, and 5 (Cohort 1, PreK–5). Threshold for adequate implementation (at the program level) = 4 (out of 4 possible points). School year 2020–21: Sum of indicators 1 (Cohort 2, 6-12), 2 (Cohort 2, 6-12), and 5 (Cohort 1, PreK–5), Threshold for adequate implementation (at the program level) = 3 (out of 3 possible points). School year 2021–22: Sum of indicators 4 and 5 (Cohort 2, 6-12). Threshold for adequate implementation (at the program level) = 2 (out of 2 possible points). School year 2022–23: Indicator 5 (Cohort 2, 6-12) score. Threshold for adequate implementation (at the program level) = 1 (out of 1 possible point).	

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at Sample Level
Key Component 2. Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites			
1. The program establishes work teams.	Program	**A) Community Collaboratory** 1 = Community Collaboratory established. 0 = Community Collaboratory not established. **B) District Implementation Teams (Grades PreK–5, Grades 6–12)** 1 = District Implementation Team established. 0 = District Implementation Team not established. **C) Site Implementation Teams** 1 = Site Implementation Teams established in 100% of implementing schools. 0 = Site Implementation Teams not established in 100% of implementing schools. **D) Technical Working Group** 1 = Technical Working Group established. 0 = Technical Working Group not established. **E) Leadership Council** 1 = Leadership Council established. 0 = Leadership Council not established.	N/A
2. Community Collaboratory meets on a regular basis	Program	1 = Community Collaboratory meets 4 times per year. 0 = Community Collaboratory meets fewer than 4 times per year.	N/A
3. District Implementation Teams meet on a regular basis	District	2= District Implementation Team meets more than 4 times per year. 1 = District Implementation Team meets 4 times per year. 0 = District Implementation Team meets fewer than 4 times per year.	1= 100% of District Implementation Teams score a 1 or 2. 0 = Less than 100% District Implementation Teams score a 1 or 2.
4. Site Implementation Teams meet on a regular basis	School	2= Site Implementation Team meets more than 3 times per year. 1 = Site Implementation Team meets 3 times per year. 0 = Site Implementation Team meets fewer than 3 times per year.	1 = 75% or more of Site Implementation Teams score a 1 or a 2. 0 = Less than 75% of Site Implementation Teams score a 1 or a 2.

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at Sample Level
5. Technical Working Group meets on a regular basis	Program	1 = Technical Working Group meets 2 times per year. 0 = Technical Working Group meets fewer than 2 times per year.	N/A
6. Leadership Council meets on a regular basis	Program	1 = Leadership Council meets 6 times per year in 2019–20 and 2020–21 or 4 times per year in 2021–22 and 2022–23. 0 = Leadership Council meets fewer than 6 times per year in 2019–20 and 2020–21 or fewer than 4 times per year in 2021–22 and 2022–23.	N/A
Key Component 2 Total Score		2019–20: Sum of scores across indicators 1A, 1B (Cohort 1, PreK–5), 1C (Cohort 1, PreK–5), 1D, 1E, and 2–6. Threshold for adequate implementation (at the program level) = 10 (out of 10 possible points). 2020–21: Sum of scores across indicator 1B (Cohort 2, 6–12) and indicators 2–6. Threshold for adequate implementation (at the program level) = 6 (out of 6 possible points). 2021–22: Sum of scores across indicator 1C (Cohort 2, 6–12) and indicators 2–6. Threshold for adequate implementation (at the program level) = 6 (out of 6 possible points). 2022–23: Sum of scores across indicators 1C (Cohort 3) and indicators 2–6. Threshold for adequate implementation (at the program level) = 6 (out of 6 possible points).	

Indicator	Unit of Measurement	Indicator Scoring at Unit Level	Indicator Scoring at Sample Level
Key Component 3. Provide professional support structures and capacity building to effectively implement the curricular units			
1. School site plans	School	2 = Plan developed and implemented. 1 = Plan developed but not implemented. 0 = Plan not developed and not implemented.	1 = 75% or more of schools score 1 or 2. 0 = Fewer than 75% of schools score 1 or 2.
2. Training for District Implementation Teams	District	1 = District Implementation Team receives 5 training sessions. 0 = District Implementation Team receives fewer than 5 training sessions.	1 = 100% of District Implementation Teams score 1. 0 = Fewer than 100% of District Implementation Teams score 1.
3. Support for Site Implementation Teams	School	1 = Site Implementation Team receives support at least 3 times during the year. 0 = Site Implementation Team does not receive support at least 3 times during the year.	1 = 75% or more of Site Implementation Teams score 1 or 2. 0 = Fewer than 75% of Site Implementation Teams score 1 or 2.
4. Site administrator training	School	1 = Site administrator participates in 3 or more training sessions per year. 0 = Site administrator participates in 2 or fewer training sessions per year.	1 = 75% or more of site administrators score 1 or 2. 0 = Fewer than 75% of site administrators score 1 or 2.
5. District-level training for teachers	District	1 = District Implementation Team provides 4 or more training sessions for teachers. 0 = District Implementation Team provides 3 or fewer training sessions for teachers.	1 = 100% of District Implementation Teams score 1. 0 = Less than 100% of District Implementation Teams score 1.
6. District executive leadership support	District	1 = District executive leadership engage with district and/or site leaders at least 6 times. 0 = District executive leadership engage with district and/or site leaders fewer than 6 times.	N/A
Key Component 3 Total Score		<u>2019–20:</u> Sum of indicators 1–5. (Indicator 6 was not created until the 2020–21 school year.) Threshold for adequate implementation (at the program level) = 5 (out of 5 possible points). <u>2020–21, 2021–22, 2022–23:</u> Sum of all indicators. Threshold for adequate implementation (at the program level) = 6 (out of 6 possible points).	

Appendix B. Fidelity of Implementation Findings

Abt measured fidelity of implementation during the four years of the study for all entities participating in the PreK–12 STEM Pathway (Exhibit B-1).

Exhibit B-1. Participants in Fidelity of Implementation Assessment, by Implementation Year

School Year	Sample	
2019–20	• PreK–5 Standards and Curriculum Team • Community Collaboratory • PreK–5 District Implementation Team (DIT) • 9 Site Implementation Teams (SITs)	• Technical Working Group (TWG) • Leadership Council • 9 schools • 9 site administrators
2020–21	• PreK–5 and grades 6–12 Standards and Curriculum Teams • Community Collaboratory • PreK–5 and grades 6–12 DITs • 9 SITs	• TWG • Leadership Council • 9 schools • 9 site administrators
2021–22	• PreK–5 and grades 6–12 Standards and Curriculum Teams • Community Collaboratory • PreK–5 and grades 6–12 DITs • 12 SITs	• TWG • Leadership Council • 12 schools • 12 site administrators
2022–23	• PreK–5 and grades 6–12 Standards and Curriculum Teams • Community Collaboratory • PreK–5 and grades 6–12 DITs • 14 SITs	• TWG • Leadership Council • 14 schools • 14 site administrators

The Community Training and Assistance Center (CTAC) and Tracy Unified School District (TUSD) implemented the PreK–12 STEM Pathway with fidelity in years 2–4 but not in year 1 (Exhibit B-2).

- In school year 2019–20, the PreK–12 STEM Pathway was not and could not be implemented with fidelity**.** The PreK–12 STEM Pathway met fidelity of implementation standards for key components 1 and 3 but did not meet fidelity requirements for key component 2. Out of the six indicators comprising key component two, only two thresholds were not met: Meetings for the Community Collaboratory and TWG were cancelled due to the COVID-19 public health emergency. Each of these reflected an intentional and strategic decision to prioritize people's health over program implementation during the COVID-19 public health emergency.
- In school years 2020–21, 2021–22, and 2022–23, the PreK–12 STEM Pathway met fidelity of implementation standards for all key components.

Exhibit B-2. Program Fidelity Results

Key Component	Fidelity Threshold Met?			
	School Year 1 2019–20	School Year 2 2020–21	School Year 3 2021–22	School Year 4 2022–23
Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science	Yes	Yes	Yes	Yes
Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites	No	Yes	Yes	Yes

Key Component	Fidelity Threshold Met?			
	School Year 1 2019–20	School Year 2 2020–21	School Year 3 2021–22	School Year 4 2022–23
Provide professional support structures and capacity building to effectively implement the curricular units	Yes	Yes	Yes	Yes

Below, we provide a more detailed description of fidelity of implementation results.

School Year 1 (2019–20)

In 2019–20, the program built curricular units and provided professional support structures and staff as intended. While all work teams were established, not all teams met as intended due to the COIVD-19 public health emergency. Below, we highlight key activities for each of the three areas (Exhibit B-3).

Exhibit B-3. School Year 1 Implementation Fidelity Findings

Key Components, Number of Indicators, Units, and Threshold				Year 1 Results *(2019–20 School Year)*		
Key Component	**Total # of Measurable Indicators That Year**	**Unit of Implementation**	**Sample-Level Threshold For Fidelity of Implementation**	**Number Of Units in Which Component Was Implemented**	**Number of Units in Which Fidelity of Component was Measured**	**Achieved Fidelity Score and Whether Program Met Sample-Level Threshold**
1. Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science	4	Program	**Sample-level component score of 4 out of 4 possible points**	1 program	1 program	Score of 4 ***Program fidelity = Yes***
2. Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites	6	4 program-level indicators 1 district-level indicator 1 school-level indicator	**Sample-level component score of 10 out of 10 possible points**	1 program 1 DIT 9 schools	1 program 1 DIT 9 schools	Score of 8 ***Program fidelity = No***
3. Provide professional support structures and capacity building to effectively implement the curricular units	5	2 district-level indicators 3 school-level indicators	**Sample-level component score of 5 out of 5 possible points**	1 DIT 9 schools	1 DIT 9 schools	Score of 5 ***Program fidelity = Yes***

Note: DIT=District Implementation Team.

Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science. In 2019–20, the program built and revised the PreK–5 curriculum.

- The program established the PreK–5 Standards and Curriculum Team to build the curriculum. The team comprised 21 members, with three team members representing each grade level from grades PreK–5.

- The program provided training for PreK–5 Standards and Curriculum team members. Across the 23 PreK–5 Standards and Curriculum Team trainings held from March 2019 to February 2020, each grade level had at least one team member present at 83% of meetings.
- The PreK–5 Standards and Curriculum Team developed the PreK–5 curricular units, consisting of 4 curricular units per grade level. Example grade 3 curricular units include “Move That Toy,” which explores how forces affect the world around us; “Growing Up,” in which students learn how living things change and grow; “Survive and Thrive,” which focuses on how the environment affects living organisms; and “Surviving Tracy Winds,” which explores how weather affects the Tracy community.
- The PreK–5 Standards and Curriculum Team met for Design Team Days in late May and early June 2020 to make general adjustments to the curricular units, as well as to adjust the curricular units in anticipation of both modified in-class and online implementation during the COVID-19 public health emergency, meeting expectations for this indicator.

Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites. In 2019–20, five sets of work teams had been established:

- The Community Collaboratory had 14 members representing various community organizations. This group met to assist with creating field experiences.
- The PreK–5 DIT had 14 members, representing district administrators and teacher leaders who supported implementation of project-based learning curricular units and pedagogy.
- Each of the nine SITs had between three and five members and was a mix of teachers and administrators at the school. These teams supported implementation of project-based learning curricular units and pedagogy.
- The TWG had three members, representing a group of experts from STEM-focused industries and PreK–12 education, that advised the Leadership Council on developments in STEM that could improve implementation.
- The Leadership Council had six members who directed and oversaw project implementation.

Work teams met throughout the year.

- The Community Collaboratory met for only three out of four of its scheduled times, with the last meeting being cancelled due to the COVID-19 public health emergency .
- The PreK–5 DIT met five times from July 2019 to March 2020. The DIT did not meet again during the school year due to the COVID-19 public health emergency.
- The SITs met between five and 10 times each. Three SITs had additional meetings planned that were cancelled due to the COVID-19 public health emergency. The focus of SIT meetings varied across teams; example content included unpacking challenges to program implementation, classroom walkthroughs, planning teacher training, and organizing STEM materials.
- The TWG met for only one of its two scheduled meetings, due to the COVID-19 public health emergency.
- The Leadership Council met eight times during the 2019–20 school year,

Provide professional support structures and capacity building to effectively implement the curricular units. The program provided supports to schools so that they could best implement the curricular units.

- All nine schools planned to implement actions within three to five areas of STEM implementation through a site plan for STEM instruction. Examples of actions that were planned and implemented within schools included supporting teachers in developing students' oral language during STEM instruction, creating a science lab for teachers and students, hosting science nights for families, and implementing rich STEM experiences.
- SITs received support from leadership teams. Eight of the nine SITs received support on three dates, and one SIT received support on nine dates. This support could come in many forms, such as professional development days or pedagogical training. Example supports included discussing data collection from walkthroughs, co-facilitating school workshops with the SIT, and helping with planning and lesson studies.
- Site administrators participated in at least three training sessions throughout the school year.
- The PreK–5 DIT provided four training sessions for teachers about the curricular units. A fifth training was planned but was cancelled due to the COVID-19 public health emergency.

The program additionally provided supports to the PreK–5 DIT.

- The PreK–5 DIT received five program training sessions throughout the year.

School Year 2 (2020–21)

In 2020–21, the program built curricular units, established work teams, and provided professional support structures and staff as intended (Exhibit B-4).

Exhibit B-4. School Year 2 Implementation Fidelity Findings

Key Components, Number of Indicators, Units, and Threshold				Year 2 Results (2020–21 School Year)		
Key Component	Total # of Measurable Indicators That Year	Unit of Implementation	Sample-Level Threshold for Fidelity of Implementation	Number of Units in Which Component was Implemented	Number of Units in Which Fidelity of Component was Measured	Achieved Fidelity Score and Whether Program met Sample-level Threshold
1. Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science	3	Program	**Sample-level component score of 3 out of 3 possible points**	1 program	1 program	Score of 3 ***Program fidelity = Yes***
2. Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites	6	4 program- level indicators 1 district level indicator 1 school level indicator	**Sample-level component score of 6 out of 6 possible points**	1 program 2 DITs 9 schools	1 program 2 DITs 9 schools	Score of 6 ***Program fidelity =Yes***
3. Provide professional support structures and capacity building to effectively implement the curricular units	6	3 district-level indicators 3 school-level indicators	**Sample-level component score of 6 out of 6 possible points**	1 DIT 9 schools	1 DIT 9 schools	Score of 6 ***Program fidelity = Yes***

Note: DIT=District Implementation Team. The 6-12 DIT was established in year 2 but was not required to meet or receive support until Cohort 2 curriculum delivery began in year 3.

Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science. In 2020–21, the program began to build the grades 6–12 curricular units and revised the PreK–5 curricular units.

- The program established the grades 6–12 Standards and Curriculum Team. The team comprised 21 lead teachers spanning grade levels and math and science content areas. Most content teams had two members, except for Grade 8 Science, Geometry and Algebra II, and Computer Science, which had one member each.
- The program trained the grades 6–12 Standards and Curriculum team. There were 11 training sessions held during the 2020–21 school year, covering content such as engineering, content progression, assessment, and project-based learning. The majority of content teams had at least one member present at all sessions.
- The grades 6–12 Standards and Curriculum Team began building the 6-12 curricular units.
- The PreK–5 Standards and Curriculum Team revised the PreK–5 curriculum multiple times during the year. District STEM leaders additionally reviewed each unit to determine the materials needed for conducting the unit in a distance learning environment.

Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites. The program established the grades 6–12 DIT, consisting of 10 members spanning grades 6–12. All other work teams established in 2019–20 continued to operate.

Each work team met throughout the year.

- The Community Collaboratory met four times.
- The PreK–5 DIT met five times.
- The nine SITs met between four and 19 times each. The focus of SIT meetings varied across teams; example content includes goal setting, integrating STEM with hybrid/distance learning, equitable access for all students, sense-making notebooks, data review, and STEM walkthroughs.
- The TWG met two times.
- The Leadership Council met 10 times.

Provide professional support structures and capacity building to effectively implement the curricular units. The program provided supports to schools so that they could best implement the curricular units.

- Schools planned to implement actions within three to five areas of STEM implementation through a site plan for STEM instruction. Examples of actions that were planned and implemented within schools include implementing sense-making notebooks, increasing engagement, integrating STEM and other subjects, high-level questioning, providing collaborative planning time, and providing professional development.
- SITs received support from leadership teams, with support provided three to 17 times during the school year. Examples of support include discussing sense-making notebooks, reviewing rubrics, conducting walkthroughs, and holding professional development about STEM storyboarding.
- Seven out of nine site administrators participated in three or more training sessions.

- The PreK–5 DIT provided four training sessions per grade level, one for each curricular unit.
- District executive leadership engaged district and/or site leaders in bimonthly STEM conversations. District leaders held four management meetings between July 2020 and May 2021 in addition to holding two to three STEM meetings with principals between November 2020 and February 2021.

The program additionally provided supports to the PreK–5 DIT.

- The PreK–5 DIT received five training sessions

School Year 3 (2021–22)

In 2021–22, the program built curricular units, established work teams, and provided professional support structures and staff as intended (Exhibit B-5).

Exhibit B-5. School Year 3 Implementation Fidelity Findings

Key Components, Number of Indicators, Units, and Threshold				Year 3 Results *(2021–22 School Year)*		
Key Component	Total # of Measurable Indicators That Year	Unit of Implementation	Sample-Level Threshold for Fidelity of Implementation	Number Of Units in Which Component was Implemented	Number Of Units in Which Fidelity Of Component was Measured	Achieved Fidelity Score and Whether Program Met Sample-Level Threshold
1. Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science	2	Program	**Sample-level component score of 2 out of 2 possible points**	1 program	1 program	Score of 2 ***Program fidelity = Yes***
2. Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites	6	4 program-level indicators 1 district- level indicator 1 school-level indicator	**Sample-level component score of 6 out of 6 possible points**	1 program 2 DITs 12 schools	1 program 2 DITs 12 schools	Score of 6 ***Program fidelity =Yes***
3. Provide professional support structures and capacity building to effectively implement the curricular units	6	3 district-level indicators 3 school-level indicators	**Sample-level component score of 6 out of 6 possible points**	2 DITs 12 schools	2 DITs 12 schools	Score of 6 ***Program fidelity = Yes***

Note: DIT=District Implementation Team.

Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science. In 2021–22, the program finished building the grades 6-12 curricular units and revised the grades 6-12 curricular units.

- The grades 6-12 Standards and Curriculum Team developed 24 total curricular units, including one fall and one spring unit in each subject area: grade 6 science, grade 6 math, grade 7 science, grade 7 math, grade 8 science, grade 8 math, biology, chemistry, physics, geometry, algebra I, and algebra II.

Example curricular units include the grade 8 math unit “Figure a Flight Plan,” in which students explore forces on a model aircraft during flight and the physics unit “Design a Crash Cart,” in which students conduct impulse calculations and engage with graphical representations of their crash cart data.

- The grades 6–12 Standards and Curriculum Team revised the curriculum. The revisions were made by each grade-level or subject-level team at the end of the school year. Deliverables included reviewing teacher feedback and providing suggested recommendations and building a “5E” (engage, explore, explain, elaborate, evaluate) exemplary lesson sequence for the first key concept in each STEM unit.

Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites. All work teams established previously continued to operate. The number of SITs increased to 12 with the addition of grades 6–12 campuses implementing the curricular units.

Each work team met throughout the year.

- The Community Collaboratory met four times.
- The DITs met a total of nine times.
- The number of SIT meetings varied, with each team meeting three to 29 times during the 2021–22 school year. The majority of teams met six to 10 times. The focus of SIT meetings varied across teams; example content included conducting walkthroughs, discussing academic discourse and real-world connections in instruction, evaluating student engagement strategies, discussing how to assist teachers with uploading artifacts, and reviewing supply needs and the budget.
- The TWG met two times.
- The Leadership Council met six times.

Provide professional support structures and capacity building to effectively implement the curricular units. The program provided supports to schools so that they could best implement the curricular units.

- Schools planned to implement actions within three to nine areas of STEM implementation through site plans for STEM instruction. Examples of actions that were planned and implemented within schools include helping teachers understand and teach the engineering design process, supporting 5E lesson planning, implementing Claims Evidence Reasoning, and using sense-making notebooks.
- SITs received between three and 25 supports during the school year from leaderships teams. This support could come in many forms, such as professional development days or pedagogical training. Example supports include Teacher on Special Assignment (TOSA) support for STEM curricular units in various grade levels; rubric calibration; co-observations of STEM instruction; and coaching about rigor, relevance, and engagement.
- Site administrators participated in five to eight training sessions each.
- The DITs each held four training sessions for teachers about curricular unit topics. The PreK–5 training sessions focused on fostering active participation in the classroom. The grades 6–12 training sessions focused on the implementation of STEM curricular units.
- District executive leadership engaged district and/or site leaders in STEM conversations. District leaders held six management meetings between October 2021 and February 2022 in addition to holding one STEM meeting with each principal.

The program additionally provided supports to the PreK–5 and grades 6-12 DITs.

- Both teams received a total of nine total training sessions.

School Year 4 (2022–23)

In 2022–23, the program built curricular units, established work teams, and provided professional support structures and staff as intended (Exhibit B-6).

Exhibit B-6. School Year 4 Implementation Fidelity Findings

Key Components, Number of Indicators, Units, and Threshold				Year 4 Results (2022–23 School Year)		
Key Component	Total # of Measurable Indicators That Year	Unit of Implementation	Sample-Level Threshold For Fidelity of Implementation	Number Of Units in Which Component was Implemented	Number Of Units in Which Fidelity of Component was Measured	Achieved Fidelity Score and Whether Program Met Sample-Level Threshold
1. Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science	1	Program	**Sample-level component score of 1 out of 1 possible point**	1 program	1 program	Score of 1 ***Program fidelity =Yes***
2. Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites	6	4 program-level indicators 1 district- level indicator 1 school-level indicator	**Sample-level component score of 6 out of 6 possible points**	1 program 2 DITs 14 schools	1 program 2 DITs 14 schools	Score of 6 ***Program fidelity =Yes***
3. Provide professional support structures and capacity building to effectively implement the curricular units	6	3 district-level indicators 3 school-level indicators	**Sample-level component score of 6 out of 6 possible points**	2 DITs 14 schools	2 DITs 14 schools	Score of 6 ***Program fidelity = Yes***

Note: DIT=District Implementation Team.

Build and refine STEM curricular units that are interdisciplinary, project-based, and focus on engineering and computer science. In 2022–23, the program revised the grades 6–12 curricular units. The grades 6–12 Standards and Curriculum Team made revisions by grade-level or subject-level team at the end of the school year.

Establish a multi-tiered system of leadership teams to guide and inform implementation of the curricular units at the school sites All previously established work teams continued to operate. The number of SITs increased to 14 with the addition of Cohort 3 schools implementing the PreK–5 curricular units.

Work teams met throughout the year.

- The Community Collaboratory met four times.
- The DITs met a total of 10 times.
- The SITs met five to 21 times each. The majority of teams met eight to 10 times. The focus of SIT meetings varied across teams; example content included discussing rigor, working on 5E lesson plans, reflection on implementation, conducting classroom visits and walkthroughs, and outlining lesson support for teachers.
- The TWG met two times.
- The Leadership Council met seven times.

Provide professional support structures and capacity building to effectively implement the curricular units. The program provided supports to schools so that they could best implement the curricular units.

- Schools planned to implement actions within three to nine areas of STEM implementation through a site plan for STEM instruction. Examples of actions that were planned and implemented within schools included reviewing the STEM curricular units, training and supporting teachers in 5E lesson design, determining teacher needs, and facilitating a carousel walk of student artifacts.
- SITs received support from leadership teams from five to 11 times each. This support could come in many forms, such as professional development days or pedagogical training. Example supports include site planning guidance, answering team questions and concerns, meetings about maximizing impact and navigating resources, coaching teachers to develop the 5Es, implementing science professional development for teachers, and providing STEM walkthrough support.
- Site administrators participated in five training sessions each.
- The DITs each held five trainings sessions for teachers about curricular unit topics. The PreK–5 training sessions focused on inquiry and student sense-making. The grades 6–12 training sessions focused standard implementation and assessing student learning to inform instruction.
- District executive leadership engaged district and/or site leaders in STEM conversations. District leaders held 13 management meetings between August 2022 and April 2023 in addition to holding one to three STEM meetings with each principal.

The program additionally provided supports to the PreK–5 and grades 6–12 DITs.

- The PreK–5 and grades 6–12 DITs received a total of 10 training sessions.

Appendix C. Impact Study Design Supplemental Information

This appendix includes supplemental information on outcome measures, the comparison school selection procedure, the analytic approach, baseline equivalence estimation, and how representativeness was examined.

Appendix C.1 Outcome Measures

In this section, we describe the outcome measures, their domains and reliability, the years in which they were available, and how they were constructed and collected.

Exhibit C-1 below summarizes the outcomes used in the evaluation. All data were publicly available, downloaded from the California Department of Education website.[17] For each outcome, although the study team collected multiple years of baseline data to maximize power, the study team established baseline equivalence (discussed in further detail below) using the year immediately prior to implementation (i.e., baseline equivalence was established in 2018–19 for Cohort 1 and in 2020–21 for Cohort 2).

The primary outcomes used in the impact analysis were English language arts (ELA), math, and science achievement. ELA and math achievement data were available for grades 3–8 and 11, and science achievement data were available for grades 5, 8, and 10–12. Each achievement outcome comprised school-level mean scores for each grade level and year. Each school contributed multiple years of ELA, math, and science scores for the grades tested, consistent with the school's grade configuration. Because the study team combined data across grades into a single confirmatory analysis for these outcomes, it converted the school-by-grade-by-year scores into standardized z-scores prior to analysis. To do this, it used the student-level statewide score means and standard deviations for the relevant testing year. For example, to calculate a school's z-score for grade 4 for school year 2018–19 from the published mean scale score for that school, the study used the student-level statewide score mean and standard deviation for the 2018–2019 school year for grade 4.

For each college readiness outcome, the California Department of Education reports the number of students in the high school graduating cohort who met the relevant criterion (passing two or more Advanced Placement (AP) exams or completing the California College Prepared requirements).[18]

[17] ELA, math, and science achievement data were available within the California Department of Education's California Assessment of Student Performance and Progress (California Department of Education, 2024a). The AP pass rate and California College Prepared outcomes are available as part of the California Department of Education's College/Career Indicator data, which are reported in the California School Dashboard (California Department of Education, 2023b).

[18] The University of California (UC) and the California State University (CSU) systems have established statewide guidelines to communicate expectations about college readiness and admission. To be considered "College Prepared" by the UC and CSU system, a student must first meet a uniform minimum set of courses required for admission as a freshman (the "A–G" course requirements). These cover courses in History/Social Studies, English, Math, Science, Languages other than English, Visual and Performing Arts, and College Preparatory Electives. To be considered College Prepared, students must complete the A–G course requirements with a grade of at least a C- (for 2017–18, 2018–19, and 2022–23) or a C (for 2019–20, 2020–21, and 2021–22), plus meet one of these additional criteria: (1) earn at least a Level 3 ("Standard Met") on the ELA or math Smarter Balanced Summative Assessment and at least a Level 2 on the other assessment; (2) complete one semester (or the equivalent) of academic or career and technical education (CTE) courses with a grade of C- or better, where college credits are awarded for each course; (3) earn a score of 3 on one AP exam or score of 4 on one International Baccalaureate exam; or (4) complete the CTE Pathway. For more information, see California Department of Education (2024c) and California Department of Education (2019).

However, it is possible that, after implementation began, the PreK–12 STEM Pathway could affect whether a student is or is not classified as being part of the graduating cohort (e.g., by affecting what courses they take, what grades they earn, and whether they are on track to graduate). As a result, to reduce bias in its outcome measurement, the study team calculated AP pass rates and California College Prepared rates by dividing the number of students who met the relevant criteria (passing two or more AP exams or satisfying California's College Prepared requirements) by the number of grade 12 students enrolled at the school, rather than by the size of the graduating cohort.

Moreover, because there are two outcomes within the college readiness domain, the study team computed the correlation of these two outcomes across all schools at the follow-up time point for which impacts were analyzed in this study (school year 2022–23). The Pearson correlation coefficient is 0.247 (p=0.125).

Exhibit C-1. Outcome Measures Used in Impact Analysis

Outcome Domain[a]	Outcome Measure	Outcome Construction	Reliability/ Validity	Baseline Measure(s) (at the school level)
Literacy Achievement	**ELA achievement**	Grade-level Smarter Balanced ELA scores for grades 3–8 and 11 in standard deviation units (z-scores) for two to three years (2020–21 through 2022–23), depending on cohort. Assessed after 2 years of exposure for confirmatory analyses and some exploratory analyses, and after three years for some exploratory analyses.	Standardized test; assumed reliable and valid	*Same as outcome*. Grade-level Smarter Balanced ELA scores for grades 3–8 and 11 in standard deviation units (z-scores)
Math Achievement	**Math achievement**	Grade-level Smarter Balanced math scores for grades 3–8 and 11 in standard deviation units (z-scores) for two to three years (2020–21 through 2022–23) depending on cohort. Assessed after two years of exposure for confirmatory analyses and some exploratory analyses, and after three years for some exploratory analyses.	Standardized test; assumed reliable and valid	*Same as outcome.* Grade-level Smarter Balanced math scores for grades 3–8 and 11 in standard deviation units (z-scores)
Science Achievement	**Science achievement**	Grade-level California Science Test (CAST) score for grades 5, 8, and one of grades 10–12 in standard deviation units (z-scores) for two to three years (2020–21 through 2022–23) depending on cohort. Assessed after two years of exposure for confirmatory analyses and some exploratory analyses, and after three years for some exploratory analyses.	Standardized test; assumed reliable and valid	Grade-level CAST score for grades 5, 8, and one of grades 10–12 in standard deviation units (z-scores)
College readiness	**AP exam pass rate[b]**	Percentage of grade 12 students who earn a score of 3 or higher (a passing score) on at least two AP exams. Assessed after two years of exposure (2021–22 and 2022–23).	Administrative data; assumed reliable and valid	*Same as outcome.* Percentage of grade 12 students who earn a score of 3 or higher on at least two AP exams
	California College Prepared rate[c]	Percentage of grade 12 students meeting the UC and CSU systems' requirements to be "prepared" for college. Assessed after two years of exposure (2021–22 and 2022–23).	Administrative data; assumed reliable and valid	*Same as outcome.* Percentage of grade 12 students meeting California College Prepared requirements

AP=Advanced Placement. CAST=California Science Test. ELA=English language arts.
[a] Outcome domains are aligned with definitions in the WWC Study Review Protocol, Version 5.0 (U.S. Department of Education, 2023).
[b] Given the available data and the fact that a given student can take multiple AP exams, the study team defined AP exam pass rate as the percentage of AP exams taken by students at the school in which the test taker earned a 3 ("qualified") or higher, based on the College Board's AP exam score definitions (College Board, n.d.). The study team defined this metric for grade 12 students only given that the College and Career Indicator data from the California Department of Education (2023b) focuses on students in the graduating high school cohort.
[c] The University of California and the California State University systems have established a uniform minimum set of courses required for admission as a freshman, which are referred to as the "A–G" course requirements. These cover courses in History/Social Studies, English, Math, Science, Languages other than English, Visual and Performing Arts, and College Preparatory Electives. The California College Prepared outcome in this study captures the percentage of students who complete the A–G course requirements with a grade of at least a C- (for 2017–18, 2018–19, and 2022–23) or a C (for 2019–20, 2020–21, and 2021–22), plus meet one of these additional criteria: (1) earn at least a Level 3 ("Standard Met") on the ELA or math Smarter Balanced Summative Assessment and at least a Level 2 on the other Assessment; (2) complete one semester (or the equivalent) of academic or career and technical education (CTE) courses with a grade of C- or better, where college credits are awarded for each course; (3) earn a score of 3 on one AP exam or score of 4 on one International Baccalaureate exam; or (4) complete the CTE Pathway. For more information, see California Department of Education (2024c) and California Department of Education (2019).

Exhibit C-2 shows the years when data were available for each outcome in each cohort. In general, the study team used baseline data dating back to the 2014–15 school year, when possible, consistent with data availability and with the outcome metric being defined consistently across years. ELA and math achievement data were available for all students and for Hispanic students, boys, and girls separately, starting in 2014–15; for those analyses, the study team therefore had five years of baseline data for Cohort 1 and six years of baseline data for Cohort 2. MLL students were only reported as a separate category starting in 2016-17, resulting in slightly fewer years of baseline data on ELA and math achievement for that subgroup. Science achievement data were only available starting in 2018–19, the year the CAST became operational, resulting in only one year of baseline data for Cohort 1 and two years of baseline data for Cohort 2. ELA, math, and science achievement data, which reflect standardized test scores, were not available for the 2019–20 school year due to test cancellations related to the COVID-19 pandemic. Finally, both college readiness outcomes, relevant to Cohort 2 only, had consistent baseline data available starting in 2017–18.[19]

Exhibit C-2. Years of Data Collection for Each Outcome Used in Impact Analysis

	Cohort	2014–15	2015-16	2016-17	2017–18	2018–19	2019–20	2020–21	2021–22	2022–23
ELA	1	x	x	x	x	x	N/A	T	T	.
	2	x	x	x	x	x	N/A	x	T	T
Math	1	x	x	x	x	x	N/A	T	T	.
	2	x	x	x	x	x	N/A	x	T	T
Science	1	.	.	.	.	x	N/A	T	T	.
	2	.	.	.	.	x	N/A	x	T	T
California College Prepared rates	2	.	.	.	x	x	x	x	T	T
AP pass rates	2	.	.	.	x	x	x	x	T	T

Key:
x=a pre-treatment year when a school-level baseline was obtained.
T=a year in which treatment schools in that cohort received the PreK–12 STEM Pathway program and comparison schools received business-as-usual.
..=a year in which data were not collected for the specified outcome and cohort.

[19] The College and Career Readiness Indicator data, which includes the AP pass rate and California College Prepared rate outcomes, became available starting in school year 2016–17. However, the study team excluded the 2016–17 baseline year for each outcome given differences in how AP test and California College Prepared outcomes were reported in 2016–17 relative to later baseline years.

N/A=Not available; test score data were not available in school year 2019–20 due to test cancellations as a result of the COVID-19 public health emergency.

Appendix C.2 Comparison School Selection Procedure

The study team selected comparison schools from districts across California using administrative baseline outcome and school demographic data from the California Department of Education. The study team excluded charter schools and alternative schools (such as special education or juvenile justice schools) from the potential comparison pool to ensure comparability with the treatment schools. In addition, the study team required that each comparison school have non-missing outcome data across all relevant grade levels (e.g., grades 3–5 for ELA achievement in Cohort 1 schools) both in the year immediately prior to the implementation (2018–19 for Cohort 1, 2020–21 for Cohort 2) and at the follow-up time points (2020–21 for two-year outcomes and 2021–22 for three-year outcomes for Cohort 1; 2022–23 for two-year outcomes for Cohort 2). This ensured that schools had complete data (and that different levels of data missingness between the treatment and comparison schools would not bias results), that the team could assess how similar the groups were on baseline characteristics, and that the schools could be included in the impact analysis.

The study team then conducted matching, using a combination of propensity score matching and exact matching, to select the comparison schools:

- Among the schools that met the eligibility criteria for available data, the study team divided the baseline outcome distribution into bands and used the school-level average value of the outcome from the baseline year immediately prior to the start of implementation (i.e., 2018–19 for Cohort 1 and 2020–21 for Cohort 2). The size and the end points of the bands were based on the distribution of treatment schools' baseline outcome values. The study team then conducted **exact matching** on school-grade configuration crossed with baseline achievement band. This allowed the team to identify comparison schools whose baseline outcomes and grade configurations were similar to those of the baseline treatment school within the relevant baseline outcome band.
- In conducting **propensity score matching**, the study team estimated propensity scores based on a combination of baseline achievement in the year immediately prior to implementation as well as school-level percentages of White, Black, and Hispanic students; of students identified as socioeconomically disadvantaged, and of English learners. If a potential comparison school could be matched to more than one treatment school, the matching model prioritized the match with the smallest linear distance. To ensure that the Tracy Unified School District (TUSD) comparison schools were included in the matched sample, the study team manually matched each TUSD comparison school to the most similar treatment school when the TUSD comparison school was not automatically matched through the matching algorithm described above. Including the TUSD comparison schools avoided an *n*=1 confound at the district level (U.S. Department of Education, 2022).

  The study team conducted the matching process separately for each cohort and for each impact analysis it planned to conduct (as per the research questions in Section 2 and analyses in Exhibit 4-2). In estimating the propensity scores, the study team used different combinations of baseline achievement and the school-level demographics listed in the previous paragraph, as needed, to achieve baseline equivalence and both baseline and follow-up representativeness.

As a result, each impact analysis had its own sample of comparison schools. For example, the study team had separate samples for ELA achievement after two years of exposure for all students, ELA achievement after three years of exposure for all students, ELA achievement after two years of exposure for MLL students, math achievement after two years of exposure for all students, and so forth.

Of note, the California Department of Education allowed districts to choose whether to administer the state standardized Smarter Balanced and California Science Test (CAST) tests in school year 2020–21

due to the COVID-19 pandemic. Many districts opted not to administer the state tests, resulting in ELA and math test score data being available for only 24% of the students in California and science achievement data being available for only 16% of students in California that year (ETS, 2023a; 2023b). The resulting large rates of data missingness for ELA, math, and science achievement in 2020–21 limited the pool of eligible comparison schools for ELA, math, and science outcome samples, as comparison schools in both cohorts needed to have non-missing data in 2020–21 (the two-year follow-up time point for Cohort 1 and baseline year for Cohort 2).

Although the team attempted to match each treatment school with as many comparison schools as possible, the team was typically able to find only two to five comparison schools for each treatment school for the ELA, math, and science achievement outcomes, depending on the sample. For the college readiness outcomes, which did not have as much missing data in 2019–20 or 2020–21, and for which there were only two treatment schools in the sample, the study team achieved a ratio of 10 comparison schools to each treatment school.

Appendix D provides the matching results and further information about the matching process.

Appendix C.3 Analytic Approach

The study team used a school-level **comparative short interrupted time series** (C-SITS) design to estimate confirmatory impacts of the intervention on our outcomes of interest. For ELA and math achievement and the college readiness outcomes, for which at least four years of baseline years were available for each outcome, the study team used the *baseline linear trend model version* of C-SITS. This model accounts for any possible trends in the outcome data in the years prior to the start of the intervention that might continue in the future. Because science achievement data was only available starting in the 2018–19 school year, the study team only had one year of baseline science data for Cohort 1 and two years of baseline science data for Cohort 2. Because one or two years of baseline data were not sufficient to establish a trend, for the science achievement analyses the study team used the *baseline mean projection model.* The two models are similar except the baseline mean projection model assumes that differences between the treatment group and comparison group do not change over time during the baseline years, similar to a difference-in-differences model (Hallberg et al., 2018).

Baseline Linear Trend Model for ELA and Math Achievement and College Readiness Outcomes

For ELA and math achievement, the impact model is a three-level model with repeated observations over years (level 1), nested within grades (level 2), and multiple grades, nested in schools (level 3). The subscripts *I, G,* and *j* represent the i^{th} time point for the G^{th} grade, in the j^{th} school. The dependent variable is the z-score at the i^{th} time point for the G^{th} grade, in the j^{th} school. The model includes random intercepts for schools (denoted as a_j^{Schs}), random intercepts for grades (denoted as r_{Gj}^{Grades}), random time slope terms for schools (denoted as a_{1j}^{Schs}) and for grades (r_{1Gj}^{Schs}), residual error term (denoted as ε_{iGj}), indicator variables for treatment school and for treatment school interacted with treatment year, fixed effects dummy variables for matching blocks, fixed effects dummy variables for matching blocks interacted with treatment year, and time. The random terms a_j^{Schs}, r_{Gj}^{Grades}, and ε_{iGj} are each assumed to be distributed normally with mean zero and variances σ^2_{Schs}, σ^2_{Grades}, and σ^2_{Years}, respectively, and are assumed to be independent of one another. The other model terms are described below.

The model below assesses impacts after three years of implementation.

$$
\begin{aligned}
Z_{ij} = \left(\beta_0 + a_{0j}^{Schs} + r_{0Gj}^{Grades}\right) + \beta_1\left(TrtSch_j\right) + (\beta_2 + a_{1j}^{Schs} + r_{1Gj}^{Grades})\left(Time_{ij}\right) \\
+\beta_3\left(TrtSch_j * Time_{ij}\right) \\
+ \beta_4\left(TrtSch_j * I_1\right) + \beta_5\left(TrtSch_j * I_2\right) + + \beta_6\left(TrtSch_j * I_3\right)
\end{aligned}
$$

$$+\sum_{m=1}^{M-1}\beta_{(6+m)}\,(MatchingBlock_m)$$

$$+\sum_{m=1}^{M}\beta_{(5+M+m)}\,(I_1 * MatchingBlock_m)$$

$$+\sum_{m=1}^{M}\beta_{(5+2M+m)}\,(I_2 * MatchingBlock_m)$$

$$+\sum_{m=1}^{M}\beta_{(5+3M+m)}\,(I_3 * MatchingBlock_m)$$

$$+\varepsilon_{iGj}$$

where,

β_0	=	the intercept, which is the comparison school mean score in pre-treatment years for schools in the omitted matching block.
a_{0j}^{Schs}	=	a random intercept term for school *j*. It is the deviation of school *j*'s intercept from the mean intercept, conditional on model covariates, distributed with mean 0 and variance τ_{00}^2.
r_{0Gj}^{Grades}	=	a random intercept term for grades within schools. It is the deviation of grade *G*'s in school *j*'s intercept from the mean intercept of school *j*, conditional on model covariates, distributed with mean 0 and variance ω_{00}^2.
β_1	=	the average difference between the intercepts of treatment and comparison schools at time zero.
$\boldsymbol{TrtSch_j}$	=	coded as 1 for treatment schools and 0 for comparison schools.
β_2	=	the pre-treatment time slope for comparison schools.
a_{1j}^{Schs}	=	a random time slope term for schools. It is the deviation of school *j*'s pre-intervention slope from the mean pre-intervention slope, conditional on model covariates, assumed distributed with mean 0 and variance τ_{11}^2.
r_{1Gj}^{Grades}	=	a random time slope term for grades. It is the deviation of grade *G*'s within school *j*'s pre-intervention slope from the mean of school *j*'s pre-intervention slope, conditional on model covariates, assumed distributed with mean 0 and variance ω_{11}^2.
$\boldsymbol{Time_{ij}}$	=	Continuous variable for the time period (school year), set to 0 for the last baseline year for each cohort (i.e., school year 2018–19 for Cohort 1 and 2019–20 for Cohort 2).
β_3	=	The average difference between the pre-treatment time slopes for treatment and comparison schools.
β_4	=	The treatment effect after the first year of exposure. No impact is expected this year, and this estimate will not be reported.

I_1	=	Indicator for the first year of exposure, equal to 1 for the first year of exposure (start-up year when impact is not expected), and 0 otherwise.
β_5	=	The treatment effect after the second year of exposure, for Cohorts 1 and 2. This is the estimate of interest in the confirmatory analyses and some of the exploratory analyses.
I_2	=	Indicator for the second year of implementation, equal to 1 for the second year of exposure and 0 otherwise. This is the period of interest for the confirmatory analyses and some of the exploratory analyses.
β_6	=	The treatment effect after the third year of exposure, for Cohort 1 only. This is an estimate of interest in some of the exploratory analyses.
I_3	=	Indicator for the third year of exposure, equal to 1 for the third year of exposure, and 0 otherwise. This is the period of interest for some of the exploratory analyses.
$\beta_{(6+m)}$	=	the m^{th} coefficient for the m^{th} matching block dummy variable.
$MatchingBlock_m$	=	An indicator variable that takes the value 1 if school was in the m^{th} of M matching blocks, and 0 otherwise. (Note: if there are M matching blocks, there will be M-1 matching block indicators.)
$\beta_{(6+M+m)}$	=	the m^{th} coefficient for the interaction between treatment year and the m^{th} matching block dummy variable.
$\boldsymbol{\beta_{(5+M+m)}}$	=	The difference between the model-projected comparison school mean for schools in Block *m* at post-treatment year 1 and the observed mean for comparison schools at that time point.
$\beta_{(5+2M+m)}$	=	The difference between the model-projected comparison school mean for schools in Block *m* at post-treatment year 2 and the observed mean for comparison schools at that time point.
$\beta_{(5+3M+m)}$	=	The difference between the model-projected comparison school mean for schools in Block *m* at post-treatment year 2 and the observed mean for comparison schools at that time point.
ε_{iGj}	=	The random error effect representing the difference between mean score at year *I* in grade *G* for school *j* and the predicted mean score for school *j*. These residual effects are assumed normally distributed with mean 0 and variance σ^2_{Years}, and are assumed to have 1st order autoregressive correlation. They are assumed independent from a_j^{Schs} and r_{Gj}^{Grades}.

In this model, the treatment effect after the second year of exposure is β_5 (the confirmatory outcome). The treatment effect after the third year of exposure is β_6 (the exploratory outcome). The study team included the third-year terms in the impact models for the exploratory analyses of effects after three years of exposure but did not include them in models focused on outcomes measured after two years of exposure. The impact model includes an adjustment for potential autocorrelation among repeated observations within grades within school. It also accounts for the fact that different matching blocks could have different numbers of schools—as occurred when TUSD comparison schools needed to be hand-matched within individual matching blocks. In addition, the model was implemented using the containment method to calculate the denominator degrees of freedom. Finally, in addition to the variables

in the impact model above, additional school-level covariates were added for any school-level demographics where the baseline differences were in the adjustment range (0.05 to 0.25 standard deviations) as per What Works Clearinghouse guidelines (U.S. Department of education, 2022). This model is similar to the models used in other studies (Somers et al., 2013; St. Clair et al., 2014).

For college readiness outcomes, a similar version of the same impact model was used, though with only two levels, given that the study team used school-level data (rather than school-grade-level data) for these outcomes. The two-level model reflects the fact that these outcomes have repeated observations over years nested within schools.

Science Assessment Impact Baseline Mean Projection Model

The study used a baseline mean projection model to estimate the impacts on science achievement. This is similar to a difference-in-differences model examining two points in time, as shown below. In this model the two-year treatment effect is reflected by β_5.

$$Z_{ij} = \left(\beta_0 + a_{0j}^{Schs}\right) + \beta_1\left(TrtSch_j\right) + \beta_5\left(TrtSch_j * I_2\right) + \sum_{m=1}^{M-1} \beta_{(6+m)}\left(MatchingBlock_m\right)$$
$$+ \sum_{m=1}^{M} \beta_{(5+2M+m)}\left(I_2 * MatchingBlock_m\right)$$
$$+\varepsilon_{iGj}$$

where model terms are similar to those described previously.

Similarly, the impact model to compute the three-year effects on science achievement is as follows. In this model the three-year treatment effect is reflected by β_6 .

$$Z_{ij} = \left(\beta_0 + a_{0j}^{Schs}\right) + \beta_1\left(TrtSch_j\right) + \beta_6\left(TrtSch_j * I_3\right) + \sum_{m=1}^{M-1} \beta_{(6+m)}\left(MatchingBlock_m\right)$$
$$+ \sum_{m=1}^{M} \beta_{(5+3M+m)}\left(I_3 * MatchingBlock_m\right)$$
$$+\varepsilon_{iGj}$$

where model terms are similar to those described previously.

For each model, in addition to the variables in the impact model above, additional covariates were added as needed to adjust for baseline differences above 0.05 standard deviations and improve precision.

Treatment of Missing Data

If a school had missing data for the whole school or specific grade levels—for example, due to cancellations as a result of COVID-19, permanent school closure, or another reason—the study team used case deletion to delete the missing observation from the analysis. To maximize comparability between treatment schools and comparison schools, the study also excluded charter schools, as well as schools classified by the California Department of Education as special education schools, opportunity schools, juvenile court schools, alternative schools of choice, continuation high schools, and youth authority facilities. Therefore, the sample of schools eligible for analysis consisted of schools with non-missing baseline and follow-up outcome data and with non-missing demographic data, which were not charter schools or one of the types of alternative schools noted above. There was no imputation of follow-up or baseline outcome data. Exhibit C-3 presents, for each outcome, the number of schools serving the grade levels at which the outcome data could be collected, as well as numbers of schools that were eligible for analysis and that were excluded from the analysis due to missing data or to school type. The study team

checked baseline equivalence and representativeness at baseline and at the relevant follow-up point to ensure that, after the missing case had been excluded from the analysis, the updated sample could still meet equivalence and representativeness requirements.

Exhibit C-3. Number of California Schools Eligible for Analysis and Excluded from Analysis

Cohort	Data Needed and Years	Grades	Outcome	Relevant Grade Levels for Outcome	CA Schools Serving Grade Levels for Outcome	CA Schools Eligible for Analysis	CA Schools Excluded Due to Missing Data or School Type*	Percentage of CA Schools with Missing Data
Cohort 1	2019 Baseline Outcome and Demographic Data 2021 Follow-up Outcome Data	PreK–5	ELA	3, 4, and 5	6,733	392	6,341	94%
			Math	3, 4, and 5		392	6,341	94%
			Science	5		392	6,341	94%
Cohort 2	2021 Baseline Outcome and Demographic Data 2023 Follow-up Outcome Data	6–12	ELA	6, 7, 8, and 11	9,266	1047	8,219	89%
			Math	6, 7, 8, and 11		1020	8,246	89%
			Science	8, and (10, 11, OR 12)		578	8,688	94%
		12	AP	12	2686	862	1,824	68%
			CA College Prepared	12		862	1,824	68%

* The study excluded schools if they were missing the outcome and demographic characteristic data necessary for analysis. The study also excluded charter schools, as well as schools classified by the California Department of Education as special education schools, opportunity schools, juvenile court schools, alternative schools of choice, continuation high schools, and youth authority facilities.

Calculation of Effect Size

For each ELA, math, and science achievement outcome, the impact estimate came directly from the respective impact model. These outcomes were already in standardized effect size units, where the standardization was relative to the standard deviation of student test scores.

Because the evaluation has two outcomes in the college readiness domain, the study team computed correlations for those outcomes to allow a domain-level average effect size to be calculated (see Exhibit C-4). The study team first calculated the correlation between those two measures in school year 2022–23, the two-year follow-up time point in which these outcomes were measured in the study (for Cohort 2). The study team also computed correlations between the baseline measure and two-year outcome measure for the years in which it was assessed in the study: 2020–21 and 2022–23 (Cohort 2’s baseline year and two-year follow-up year). All correlations were computed using two samples: (1) among all schools in California with college readiness data in 2021 and 2023, and (2) the analytic sample for the relevant outcomes. Due to the small sample sizes in the analytic sample, the correlations using analytic sample are not as high as those calculated using the larger sample.

Exhibit C-4. Correlations for College Readiness Outcomes

Outcome	All Schools with College Readiness Data in 2021 and 2023	Analytic Sample
Baseline and 2-year follow-up AP pass rates (2020–21 and 2022–23)	0.952***	.883***
Baseline and 2-year follow-up California College Prepared rates (2020–21 and 2022–23)	0.867***	.434*
2-year follow-up AP completion rates and California College Prepared rates (2022–23)	0.691***	.247

***: p<.001; *p<.05

Appendix C.4 Baseline Equivalence Estimation

For each outcome measure, the study team assessed baseline equivalence in the year immediately prior to the start of the PreK–12 STEM Pathway implementation (2018–19 for Cohort 1, 2020–21 for Cohort 2). The study team assessed baseline equivalence at the school level, which represented the average performance of an earlier cohort of students in the same grades and schools on the same measure in the year immediately prior to the start of the PreK–12 STEM Pathway program. This model is a two-level model with multiple grades (level 1) nested within schools (level 2).

$$Z_{Gj} = \beta_0 + \beta_1 T_j + \sum_{m=1}^{M-1} \beta_{(1+m)}(MatchingBlock_m) + \mu_j^{Schs}$$

Note that the coefficient β_1 represents the difference between the baseline mean scores of treatment and comparison schools. For each achievement outcome, the estimate of the difference between the treatment and comparison groups came directly from the respective regression model (as shown above) and was already in standardized effect size units. For college readiness outcomes, which were dichotomous, the study team calculated Cox's Index to compute a standardized difference in assessing baseline equivalence.

Even though it was not required because baseline outcome measures were available, the study team examined and reported on baseline equivalence on school composition with respect to race/ethnicity, socioeconomic disadvantage (proxied by Free or Reduced-Price Lunch percentage), English learner status, and gender. The study team assessed baseline equivalence for all samples, both all students and the student subgroup samples (MLL students, Hispanic students, boys, and girls). The results are presented in Appendix D.

Appendix C.5 Representativeness

Finally, one possible source of bias can occur if the students in treatment schools with outcome data are much more, or much less, representative of their schools than is true for comparison schools. For this study, the study team used administrative data from state tests and other state-collected measures. Non-participation in the state assessments and other state-collected outcome data has not been a widespread issue historically. As such, demonstrating that students in the analytic sample are representative of schools is not explicitly required for administrative data under What Works Clearinghouse Standards, Version 5.0 (U.S. Department of Education, 2022).

However, under What Works Clearinghouse (WWC) Version 5.0 standards, WWC reviewers have greater discretion related to scrutiny of the representativeness of administrative data,[20] and COVID-19 disrupted standardized testing in school year 2020–21 (the second year of exposure for Cohort 1, and final baseline year for Cohort 2), resulting in a high level of non-response in the state administrative data. For example, only 24% of students eligible to take the Smarter Balanced ELA and math tests in 2020–21 and only 16% of students eligible to take the CAST in 2020–21 had data (ETS, 2023a; 2023b). Thus, to increase the rigor of the study design, the study team assessed representativeness, both for the outcome year and for the baseline year immediately prior to the start of exposure.

To assess representativeness, the study team collected data on the number of students in the sample schools who were enrolled in the relevant grades at the time of assessment and the number of students in the schools contributing data, both overall and for the treatment and comparison groups separately. The study team assessed this both in the baseline year prior to exposure and at the follow-up time point, for each impact analysis. As a measure of lack of representativeness, for each analysis, the study team reported the student non-response rate; that is, 100% minus the ratio of the number of students with non-missing outcome data divided by the number of students expected to have outcome data.

The study met the What Works Clearinghouse standards for representativeness, demonstrating that individuals in the analytic sample were representative of clusters. Appendix D provides the results.

---

[20] As per the WWC Standards, "The WWC assumes that analyses of administrative data satisfy the representativeness requirement, unless review team leadership concludes that patterns of missing administrative data have a high risk of differing across intervention and comparison groups." (U.S. Department of Education, 2022).

Appendix D. Baseline Equivalence and Representativeness

This appendix provides the results of the baseline equivalence and representativeness assessments.

Appendix D.1 Baseline Equivalence

The study team examined baseline equivalence for each outcome using the matched schools from Cohorts 1 and 2.

Exhibit D-1 shows the results of baseline equivalence testing for the analytic samples for ELA, math, and science achievement after two years of exposure, the confirmatory analyses. The results of baseline equivalence testing for ELA, math, and science achievement after three years of exposure are presented in Exhibit D-2. The baseline outcomes are presented at the unit of school-grades; school-level demographics are presented at the unit of the school. School demographics assessed include the average percentage of students in the school who are White, Black, Hispanic, eligible for Free and Reduced-Price Lunch (as a proxy for socioeconomic disadvantage), English learners (EL), or girls. The exhibits also include the results for grade configuration (on which the study team required exact matches between comparison and treatment schools).

For each matched sample, the treatment and comparison schools were equivalent on the average school-grade-level score (standardized difference less than .05). The between-group differences on school-level covariates typically were either equivalent or within the adjustment range (standardized difference between 0.05 and 0.25). There were a small number of exceptions for the percentages of Black students across the entire school that are slightly above the adjustment range for some samples for specific subpopulations (e.g., ELA achievement for MLL students). These slightly higher standardized differences likely reflect the low percentage of Black students in the treatment schools (3%). As noted in Section 4 and Appendix C, the study team controlled for any demographic variables with differences above 0.05 standard deviations in the impact model.

Exhibit D-1. Baseline Equivalence Test Results for Each Matched Sample for Two-Year Follow-Up Analysis in Cohorts 1 and 2 (All Students)

Outcome	Variable	Comparison Observations	Treatment Observations	Unadjusted Mean for Comparison Schools	Adjusted Mean for Treatment Schools	Standardized Difference
ELA achievement	Average school-grade-level ELA (Smarter Balanced) score (z-score)	95[a]	41[b]	-0.22	-0.22	0.0048
	% White students in school	33	12	0.22	0.18	-0.1673
	% Black students in school	33	12	0.03	0.05	0.2076
	% Hispanic students in school	33	12	0.58	0.56	-0.0401
	% of FRPL students in school	33	12	0.63	0.58	-0.1124
	% of ELs in school	33	12	0.27	0.32	0.1560
	% of girls in school	33	12	0.48	0.49	0.0065
	% schools serving grades K–5	33	12	0.65	0.65	0.0000
	% schools serving grades K–8	33	12	0.40	0.40	0.0000
	% schools serving grades 6–8	33	12	0.20	0.20	0.0000
	% schools serving grades 9–12	33	12	0.33	0.33	0.0000

Outcome	Variable	Comparison Observations	Treatment Observations	Unadjusted Mean for Comparison Schools	Adjusted Mean for Treatment Schools	Standardized Difference
Math achievement	Average school-grade-level math (Smarter Balanced) score (z-score)	92[c]	41[b]	-0.24	-0.26	-0.0245
	% White students in school	32	12	0.22	0.18	-0.1552
	% Black students in school	32	12	0.03	0.05	0.2393
	% Hispanic students in school	32	12	0.61	0.56	-0.1248
	% of FRPL students in school	32	12	0.64	0.58	-0.1342
	% of ELs in school	32	12	0.26	0.32	0.1818
	% of girls in school	32	12	0.49	0.49	-0.0219
	% schools serving grades K–5	32	12	0.68	0.68	0.0000
	% schools serving grades K–8	32	12	0.38	0.38	-0.0000
	% schools serving grades 6–8	32	12	0.20	0.20	0.0000
	% schools serving grades 9–12	32	12	0.33	0.33	0.0000
Science achievement	Average school-grade-level science (CAST) score (z-score)	53[d]	15[b]	-0.16	-0.17	-0.0052
	% White students in school	45	12	0.22	0.18	-0.1820
	% Black students in school	45	12	0.04	0.05	0.0447
	% Hispanic students in school	45	12	0.54	0.56	0.0439
	% of FRPL students in school	45	12	0.65	0.59	-0.1549
	% of ELs in school	45	12	0.24	0.32	0.2305
	% of girls in school	45	12	0.49	0.48	-0.0097
	% schools serving grades K–5	45	12	0.64	0.64	-0.0000
	% schools serving grades K–8	45	12	0.41	0.41	0.0000
	% schools serving grades 6–8	45	12	0.19	0.19	0.0000
	% schools serving grades 9–12	45	12	0.33	0.33	0.0000

CAST=California Science Test. EL=English learner. ELA=English language arts. FRPL=Free or Reduced-Price Lunch.
[a] Represents the number of grade- level observations across 33 schools.
[b] Represents the number of grade-level observations across 12 schools.
[c] Represents the number of grade-level observations across 32 schools.
[d] This represents the number of school-grade-level observations from 45 schools.

Exhibit D-2. Baseline Equivalence Test Results for the Three-Year Follow-Up Sample with All Students in Cohort 1

Sample	Variable	Comparison Observations	Treatment Observations	Unadjusted Mean for Comparison Schools	Adjusted Mean for Treatment Schools	Standardized Difference
ELA achievement	Average school-grade-level ELA (Smarter Balanced) score (z-score)	60[a]	27[b]	-0.25	-0.25	0.0050
	% White students in school	20	9	0.23	0.18	-0.1973
	% Black students in school	20	9	0.04	0.04	0.0836
	% Hispanic students in school	20	9	0.57	0.58	0.0128
	% of FRPL students in school	20	9	0.67	0.63	-0.1056
	% of ELs in school	20	9	0.31	0.37	0.1490
	% of girls in school	20	9	0.49	0.48	-0.0100
	% schools serving grades K–5	20	9	0.65	0.65	0.0000
	% schools serving grades K–8	20	9	0.35	0.35	-0.0000
Math achievement	Average school-grade-level math (Smarter Balanced) score (z-score)	51[c]	27[b]	-0.24	-0.28	-0.0352
	% White students in school	17	9	0.21	0.17	-0.1760
	% Black students in school	17	9	0.03	0.05	0.1851
	% Hispanic students in school	17	9	0.59	0.58	-0.0241
	% of FRPL students in school	17	9	0.71	0.64	-0.1870
	% of ELs in school	17	9	0.31	0.38	0.1759
	% of girls in school	17	9	0.50	0.48	-0.0392
	% schools serving grades K–5	17	9	0.76	0.76	-0.0000
	% schools serving grades K–8	17	9	0.24	0.24	0.0000
Science achievement	Average school-grade-level science (CAST) score (z-score)	26	9	-0.23	-0.24	-0.0123
	% White students in school	26	9	0.21	0.17	-0.1611
	% Black students in school	26	9	0.05	0.05	0.0253
	% Hispanic students in school	26	9	0.55	0.59	0.0834
	% of FRPL students in school	26	9	0.70	0.63	-0.1806
	% of ELs in school	26	9	0.28	0.38	0.2492
	% of girls in school	26	9	0.49	0.48	-0.0113
	% schools serving grades K–5	26	9	0.69	0.69	0.0000
	% schools serving grades K–8	26	9	0.31	0.31	0.0000

CAST=California Science Test. EL=English learner. ELA=English language arts. FRPL=Free or Reduced-Price Lunch.
[a] This represents the number of school-grade-level observations from 20 schools.
[b] This represents the number of school-grade-level observations from 9 schools.
[c] This represents the number of school-grade-level observations from 17 schools.

The same procedure was conducted to examine the baseline equivalence for ELs, boys, girls, and Hispanic student subgroups, respectively, and similar results were found. Any school covariates with a standardized difference greater than 0.05 were included in the impact model to adjust for baseline differences. See Exhibits D-3 to D-6 for results.

Exhibit D-3. Baseline Equivalence Test Results for MLL Students for Two-Year Follow-Up Analysis in Cohorts 1 and 2

Sample	Variable	Comparison Observations	Treatment Observations	Unadjusted Mean for Comparison Schools	Adjusted Mean for Treatment Schools	Standardized Difference
ELA achievement	Average school-grade-level ELA (Smarter Balanced) score (z-score): MLL students subgroup	88[a]	41[b]	-0.33	-0.34	-0.0162
	% White students in school	33	12	0.21	0.18	-0.1063
	% Black students in school	33	12	0.03	0.05	0.3076
	% Hispanic students in school	33	12	0.60	0.56	-0.0975
	% of FRPL students in school	33	12	0.62	0.59	-0.0630
	% of ELs in school	33	12	0.27	0.32	0.1531
	% of girls in school	33	12	0.49	0.48	-0.0106
	% schools serving grades K–5	33	12	0.63	0.63	0.0000
	% schools serving grades K–8	33	12	0.41	0.41	0.0000
	% schools serving grades 6–8	33	12	0.20	0.20	0.0000
	% schools serving grades 9–12	33	12	0.33	0.33	0.0000
Math achievement	Average school-grade-level math (Smarter Balanced) score (z-score): MLL students subgroup	93[c]	41[b]	-0.32	-0.36	-0.0379
	% White students in school	34	12	0.23	0.18	-0.1642
	% Black students in school	34	12	0.04	0.05	0.1171
	% Hispanic students in school	34	12	0.56	0.55	-0.0211
	% of FRPL students in school	34	12	0.63	0.59	-0.1005
	% of ELs in school	34	12	0.31	0.32	0.0264
	% of girls in school	34	12	0.48	0.48	0.0042
	% schools serving grades K–5	34	12	0.65	0.65	0.0000
	% schools serving grades K–8	34	12	0.40	0.40	0.0000
	% schools serving grades 6–8	34	12	0.20	0.20	0.0000
	% schools serving grades 9–12	34	12	0.33	0.33	0.0000
Science achievement	Average school-grade-level science (CAST) score (z-score): MLL students subgroup	38[a]	15[b]	-0.28	-0.27	0.0099
	% White students in school	33	12	0.22	0.19	-0.1338
	% Black students in school	33	12	0.03	0.05	0.2413
	% Hispanic students in school	33	12	0.58	0.55	-0.0756
	% of FRPL students in school	33	12	0.62	0.58	-0.1007
	% of ELs in school	33	12	0.30	0.32	0.0573
	% of girls in school	33	12	0.49	0.48	-0.0007
	% schools serving grades K–5	33	12	0.65	0.65	0.0000
	% schools serving grades K–8	33	12	0.40	0.40	-0.0000
	% schools serving grades 6–8	33	12	0.20	0.20	0.0000
	% schools serving grades 9–12	33	12	0.33	0.33	0.0000

CAST=California Science Test. EL=English learner. ELA=English language arts. FRPL=Free or Reduced-Price Lunch.
[a] This represents the number of school-grade-level observations from 33 schools.
[b] This represents the number of school-grade-level observations from 12 schools.
[c] This represents the number of school-grade-level observations from 34 schools.

Exhibit D-4. Baseline Equivalence Test Results for Boys for Two-Year Follow-Up Analysis in Cohorts 1 and 2

Sample	Variable	Comparison Observations	Treatment Observations	Unadjusted Mean for Comparison Schools	Adjusted Mean for Treatment Schools	Standardized Difference
ELA achievement	Average school-grade-level (Smarter Balanced) score (z-score): boys subgroup	191[a]	41[b]	-0.35	-0.35	0.0007
	% White students in school	67	12	0.20	0.18	-0.1002
	% Black students in school	67	12	0.03	0.05	0.3097
	% Hispanic students in school	67	12	0.63	0.56	-0.1831
	% of FRPL students in school	67	12	0.67	0.58	-0.2064
	% of ELs in school	67	12	0.28	0.31	0.1037
	% of girls in school	67	12	0.49	0.49	-0.0032
	% schools serving grades K–5	67	12	0.66	0.66	0.0000
	% schools serving grades K–8	67	12	0.41	0.41	0.0000
	% schools serving grades 6–8	67	12	0.18	0.18	0.0000
	% schools serving grades 9–12	67	12	0.33	0.33	0.0000
Math achievement	Average school-grade-level math (Smarter Balanced) score (z-score): boys subgroup	106[c]	41[b]	-0.25	-0.26	-0.0060
	% White students in school	39	12	0.21	0.18	-0.1054
	% Black students in school	39	12	0.03	0.05	0.2428
	% Hispanic students in school	39	12	0.58	0.55	-0.0756
	% of FRPL students in school	39	12	0.66	0.57	-0.2206
	% of ELs in school	39	12	0.26	0.31	0.1298
	% of girls in school	39	12	0.50	0.49	-0.0238
	% schools serving grades K–5	39	12	0.65	0.65	0.0000
	% schools serving grades K–8	39	12	0.43	0.43	0.0000
	% schools serving grades 6–8	39	12	0.15	0.15	0.0000
	% schools serving grades 9–12	39	12	0.35	0.35	0.0000
Science achievement	Average school-grade-level science (CAST) score (z-score): boys subgroup	37[d]	15[b]	-0.21	-0.22	-0.0082
	% White students in school	32	12	0.17	0.18	0.0274
	% Black students in school	32	12	0.03	0.05	0.2712
	% Hispanic students in school	32	12	0.59	0.56	-0.0857
	% of FRPL students in school	32	12	0.64	0.59	-0.1315
	% of ELs in school	32	12	0.33	0.32	-0.0161
	% of girls in school	32	12	0.50	0.49	-0.0284
	% schools serving grades K–5	32	12	0.65	0.65	0.0000
	% schools serving grades K–8	32	12	0.38	0.38	-0.0000
	% schools serving grades 6–8	32	12	0.21	0.21	-0.0000
	% schools serving grades 9–12	32	12	0.36	0.36	0.0000

CAST=California Science Test. EL=English learner. ELA=English language arts. FRPL=Free or Reduced-Price Lunch.
[a] This represents the number of school-grade-level observations from 67 schools.
[b] This represents the number of school-grade-level observations from 12 schools.
[c] This represents the number of school-grade-level observations from 39 schools.
[d] This represents the number of school-grade-level observations from 32 schools.

Exhibit D-5. Baseline Equivalence Test Results for Girls for Two-Year Follow-Up Analysis in Cohorts 1 and 2

Sample	Variable	Comparison Observations	Treatment Observations	Unadjusted Mean for Comparison Schools	Adjusted Mean for Treatment Schools	Standardized Difference
ELA achievement	Average school-grade-level ELA (Smarter Balanced) score (z-score): girls subgroup	116[a]	41[b]	-0.11	-0.10	0.0097
	% White students in school	40	12	0.21	0.18	-0.1007
	% Black students in school	40	12	0.03	0.05	0.2943
	% Hispanic students in school	40	12	0.62	0.56	-0.1345
	% of FRPL students in school	40	12	0.65	0.60	-0.1469
	% of ELs in school	40	12	0.29	0.33	0.0984
	% of girls in school	40	12	0.49	0.48	-0.0061
	% schools serving grades K–5	40	12	0.64	0.64	-0.0000
	% schools serving grades K–8	40	12	0.40	0.40	0.0000
	% schools serving grades 6–8	40	12	0.14	0.14	0.0000
	% schools serving grades 9–12	40	12	0.36	0.36	0.0000
Math achievement	Average school-grade-level math (Smarter Balanced) score (z-score): girls subgroup	127[c]	41[b]	-0.25	-0.24	0.0065
	% White students in school	46	12	0.20	0.18	-0.0835
	% Black students in school	46	12	0.03	0.05	0.2983
	% Hispanic students in school	46	12	0.64	0.56	-0.1995
	% of FRPL students in school	46	12	0.68	0.59	-0.2237
	% of ELs in school	46	12	0.29	0.32	0.0955
	% of girls in school	46	12	0.49	0.48	-0.0023
	% schools serving grades K–5	46	12	0.64	0.64	-0.0000
	% schools serving grades K–8	46	12	0.40	0.40	0.0000
	% schools serving grades 6–8	46	12	0.16	0.16	0.0000
	% schools serving grades 9–12	46	12	0.37	0.37	0.0000
Science achievement	Average school-grade-level science (CAST) score (z-score): girls subgroup	43[d]	15[b]	-0.12	-0.11	0.0177
	% White students in school	37	12	0.17	0.18	0.0197
	% Black students in school	37	12	0.04	0.05	0.1713
	% Hispanic students in school	37	12	0.61	0.55	-0.1351
	% of FRPL students in school	37	12	0.64	0.58	-0.1654
	% of ELs in school	37	12	0.28	0.30	0.0735
	% of girls in school	37	12	0.49	0.49	-0.0134
	% schools serving grades K–5	37	12	0.63	0.63	0.0000
	% schools serving grades K–8	37	12	0.44	0.44	-0.0000
	% schools serving grades 6–8	37	12	0.15	0.15	-0.0000
	% schools serving grades 9–12	37	12	0.35	0.35	0.0000

CAST=California Science Test. EL=English learner. ELA=English language arts. FRPL=Free or Reduced-Price Lunch.
[a] This represents the number of school-grade-level observations from 40 schools.
[b] This represents the number of school-grade-level observations from 12 schools.
[c] This represents the number of school-grade-level observations from 46 schools.
[d] This represents the number of school-grade-level observations from 37 schools.

Exhibit D-6. Baseline Equivalence Test Results for Hispanic Students for Two-Year Follow-Up Analysis in Cohorts 1 and 2

Sample	Variable	Comparison Observations	Treatment Observations	Unadjusted Mean for Comparison Schools	Adjusted Mean for Treatment Schools	Standardized Difference
ELA achievement	Average school-grade-level ELA (Smarter Balanced) score (z-score): Hispanic subgroup	134[a]	41[b]	-0.31	-0.31	0.0057
	% White students in school	51	12	0.20	0.18	-0.0881
	% Black students in school	51	12	0.03	0.05	0.2365
	% Hispanic students in school	51	12	0.58	0.55	-0.0610
	% of FRPL students in school	51	12	0.56	0.58	0.0450
	% of ELs in school	51	12	0.25	0.30	0.1573
	% of girls in school	51	12	0.49	0.48	-0.0141
	% schools serving grades K–5	51	12	0.68	0.68	0.0000
	% schools serving grades K–8	51	12	0.42	0.42	0.0000
	% schools serving grades 6–8	51	12	0.18	0.18	0.0000
	% schools serving grades 9–12	51	12	0.33	0.33	0.0000
Math achievement	Average school-grade-level math (Smarter Balanced) score (z-score): Hispanic subgroup	89[c]	41[b]	-0.38	-0.38	0.0043
	% White students in school	32	12	0.20	0.18	-0.0768
	% Black students in school	32	12	0.03	0.05	0.2511
	% Hispanic students in school	32	12	0.59	0.55	-0.0992
	% of FRPL students in school	32	12	0.60	0.59	-0.0455
	% of ELs in school	32	12	0.30	0.32	0.0529
	% of girls in school	32	12	0.49	0.48	-0.0028
	% schools serving grades K–5	32	12	0.65	0.65	0.0000
	% schools serving grades K–8	32	12	0.39	0.39	0.0000
	% schools serving grades 6–8	32	12	0.15	0.15	0.0000
	% schools serving grades 9–12	32	12	0.38	0.38	0.0000
Science achievement	Average school-grade-level science (CAST) score (z-score): Hispanic subgroup	45[d]	15[b]	-0.29	-0.28	0.0108
	% White students in school	38	12	0.18	0.18	0.0062
	% Black students in school	38	12	0.04	0.05	0.1555
	% Hispanic students in school	38	12	0.58	0.55	-0.0673
	% of FRPL students in school	38	12	0.61	0.59	-0.0596
	% of ELs in school	38	12	0.28	0.31	0.0837
	% of girls in school	38	12	0.49	0.49	-0.0092
	% schools serving grades K–5	38	12	0.65	0.65	0.0000
	% schools serving grades K–8	38	12	0.43	0.43	0.0000
	% schools serving grades 6–8	38	12	0.15	0.15	-0.0000
	% schools serving grades 9–12	38	12	0.35	0.35	0.0000

CAST=California Science Test. EL=English learner. ELA=English language arts. FRPL=Free or Reduced-Price Lunch.
[a] This represents the number of school-grade-level observations from 51 schools.
[b] This represents the number of school-grade-level observations from 12 schools.
[c] This represents the number of school-grade-level observations from 32 schools.
[d] This represents the number of school-grade-level observations from 38 schools.

Exhibit D-7 presents the baseline equivalence testing results for the two college readiness outcomes after two years of exposure.

Exhibit D-7. Baseline Equivalence Test Results for the Two-Year Follow-Up Analysis Sample for College Readiness Outcomes for Cohort 2

Sample	Variable	Comparison Observations	Treatment Observations	Unadjusted Mean for Comparison Schools	Adjusted Mean for Treatment Schools	Standardized Difference
AP pass rates	School-level % of one grade 12 students passing 2 or more AP tests	21	2	0.05	0.05	-0.0251
	% White students in school	21	2	0.23	0.20	-0.0892
	% Black students in school	21	2	0.04	0.06	0.1871
	% Hispanic students in school	21	2	0.60	0.52	-0.1890
	% of FRPL students in school	21	2	0.50	0.46	-0.0953
	% of ELs in school	21	2	0.16	0.17	0.0420
	% of girls in school	21	2	0.48	0.48	0.0008
	% of schools serving grades K–8	21	2	0.00	0.00	-
	% of schools serving grades 6–8	21	2	0.00	0.00	-
	% of schools serving grades 9–12	21	2	1.00	1.00	-
California College Prepared rates	School-level % of grade 12 students completing California College Prepared requirements	20	2	0.33	0.33	0.0102
	% White students in school	20	2	0.17	0.21	0.1251
	% Black students in school	20	2	0.05	0.06	0.1290
	% Hispanic students in school	20	2	0.60	0.52	-0.1893
	% of FRPL students in school	20	2	0.50	0.46	-0.0987
	% of ELs in school	20	2	0.17	0.17	0.0130
	% of girls in school	20	2	0.48	0.48	-0.0111
	% of schools serving grades K–8	20	2	0.00	0.00	-
	% of schools serving grades 6–8	20	2	0.00	0.00	-
	% of schools serving grades 9–12	20	2	1.00	1.00	-

AP=Advanced Placement. EL=English learner. FRPL=Free or Reduced-Price Lunch.

Appendix D.2 Representativeness

The study team evaluated representativeness using the baseline sample and follow-up samples for each achievement outcome. See Exhibits D-8 and D-9 for results for the full sample. The WWC Standards Version 5.0 Handbook (U.S. Department of Education, 2022) provides guidelines for attrition. Using the optimistic boundary in this study to determine representativeness, the study team found that all samples were representative of the students enrolled in the corresponding year. Exhibit D-8 shows the representativeness for ELA, math, and science achievement for all students after two years of exposure, and Exhibit D-9 shows representativeness after three years of exposure.

Exhibit D-8. Baseline and Follow-Up Representativeness for ELA, Math, and Science Achievement after Two Years of Exposure (All Students)

Measure	Overall	Treatment Group	Comparison Group	Difference in Non-Response Rate (Treatment–Comparison)
Baseline sample				
Student non-response rate for ELA achievement (%)	8.61	5.52	10.08	-4.56
Student non-response rate for math achievement (%)	8.58	5.86	10.09	-4.23
Student non-response rate for science achievement (%)	10.95	6.78	12.24	-5.47
Follow-up sample				
Student non-response rate for ELA achievement (%)	4.32	3.09	4.90	-1.81
Student non-response rate for math achievement (%)	4.96	2.86	6.13	-3.27
Student non-response rate for science achievement (%)	4.04	2.53	4.74	-2.21

ELA=English language arts.

Exhibit D-9. Baseline and Follow-Up Representativeness for ELA, Math, and Science Achievement after Three Years of Exposure (All Students)

Measure	Overall	Treatment	Comparison	Difference in Non-Response Rate (Treatment–Comparison)
Baseline				
Student non-response rate for ELA achievement (%)	1.59	1.81	1.47	0.3
Student non-response rate for math achievement (%)	1.40	1.17	1.55	-0.4
Student non-response rate for science achievement (%)	1.25	2.06	0.91	1.1
Follow-up sample				
Student non-response rate for ELA achievement (%)	2.42	2.24	2.52	-0.28
Student non-response rate for math achievement (%)	1.58	1.62	1.55	0.06
Student non-response rate for science achievement (%)	2.19	2.14	2.21	-0.07

ELA=English language arts

Representativeness in the baseline sample and follow-up sample were also examined for each subgroup. These samples were found to be representative of the students enrolled in the schools. Exhibit D-10 summarizes the representativeness results in the four subgroups.

Exhibit D-10. Baseline Representativeness and Follow-Up Representativeness for Each Matched Subgroup Sample after Two Years of Exposure

Subgroup	Measure	Overall	Treatment Group	Comparison Group	Difference in Non-response Rate (Treatment–Comparison)
Baseline sample					
MLL students	Student non-response rate for ELA achievement (%)	8.89	4.94	11.18	-6.24
	Student non-response rate for math achievement (%)	8.25	4.93	10.03	-5.09
	Student non-response rate for science achievement (%)	7.90	5.04	9.42	-4.39
Boys	Student non-response rate for ELA achievement (%)	9.31	5.85	10.15	-4.29
	Student non-response rate for math achievement (%)	10.15	5.93	12.01	-6.08
	Student non-response rate for science achievement (%)	13.01	7.98	15.10	-7.13
Girls	Student non-response rate for ELA achievement (%)	8.74	5.13	10.33	-5.19
	Student non-response rate for math achievement (%)	10.50	5.75	12.18	-6.43
	Student non-response rate for science achievement (%)	11.33	5.57	13.16	-7.60
Hispanic students	Student non-response rate for ELA achievement (%)	13.51	5.49	15.85	-10.36
	Student non-response rate for math achievement (%)	8.86	5.77	10.44	-4.67
	Student non-response rate for science achievement (%)	10.94	6.75	12.24	-5.49
Follow-up sample					
MLL students	Student non-response rate for ELA achievement (%)	4.15	2.74	4.98	-2.25
	Student non-response rate for math achievement (%)	3.62	1.99	4.53	-2.53
	Student non-response rate for science achievement (%)	4.18	2.20	5.86	-3.66
Boys	Student non-response rate for ELA achievement (%)	5.08	3.25	5.52	-2.26
	Student non-response rate for math achievement (%)	4.63	3.02	5.32	-2.30
	Student non-response rate for science achievement (%)	4.58	3.01	5.66	-2.65
Girls	Student non-response rate for ELA achievement (%)	5.28	2.60	6.46	-3.86
	Student non-response rate for math achievement (%)	4.89	2.39	5.77	-3.38
	Student non-response rate for science achievement (%)	3.24	1.89	3.87	-1.98

Subgroup	Measure	Overall	Treatment Group	Comparison Group	Difference in Non-response Rate (Treatment–Comparison)
Hispanic students	Student non-response rate for ELA achievement (%)	5.27	2.90	5.95	-3.05
	Student non-response rate for math achievement (%)	5.83	2.77	7.32	-4.55
	Student non-response rate for science achievement (%)	3.82	2.50	4.48	-1.99

ELA=English language arts. MLL=Multilingual learner.

Representativeness was not assessed for the college readiness outcomes. Given that the denominator for both of those outcomes is grade 12 students enrolled, those samples by definition would be considered representative.

Appendix E. Impact Study Findings Tables

This appendix provides additional details about the impact findings.

The impact of the program on students after two years of exposure was examined for all students in both Cohorts 1 and 2. Appendix Exhibit E-1 presents the impact estimates on all academic outcomes for all students in both cohorts after two years of exposure, the confirmatory analyses of the study. None of these impacts is statistically significant at the 0.05 level. For sensitivity analyses, the study team also ran a baseline mean projection model for ELA and math achievement (rather than a linear trend model) and also ran a linear trend model for science (rather than the baseline mean projection model). The results are robust to the model type (i.e., the evaluation still did not detect impacts when the alternate type of model was used for each outcome).

Exhibit E-1. Impact of PreK–12 STEM Pathway on Academic Achievement Outcomes after Two Years of Exposure for All Students in Cohorts 1 and 2

Outcome	Treatment Sample Size		Comparison Sample Size		Adjusted Treatment Group Mean	Unadjusted Comparison Group Mean	Impact	Standard Error	p-Value
	Schools	School-Grades	Schools	School-Grades					
ELA	12	41	33	95	-0.29	-0.20	-0.09	0.05	0.080
Math	12	41	32	92	-0.30	-0.24	-0.07	0.05	0.192
Science	12	15	45	53	-0.20	-0.21	0.01	0.06	0.845

ELA=English language arts.

As described in Section 4, no significant effects were detected on academic achievement after three years of exposure for Cohort 1 or on college readiness outcomes for Cohort 2 (Exhibits E-2 and E-3).

Exhibit E-2. Impact of PreK–12 STEM Pathway on Academic Achievement Outcomes after Three Years of Exposure for Elementary School Students (Cohort 1 Only)

Outcome	Treatment Sample Size		Comparison Sample Size		Adjusted Treatment Group Mean	Unadjusted Comparison Group Mean	Impact	Standard Error	p-Value
	Schools	School-Grades	Schools	School-Grades					
ELA	9	27	20	60	-0.36	-0.23	-0.13	0.08	0.090
Math	9	27	17	51	-0.28	-0.24	-0.05	0.07	0.494
Science	9	9	26	26	-0.30	-0.22	-0.08	0.08	0.318

ELA=English language arts.

Exhibit E-3. Impact of PreK–12 STEM Pathway on College Readiness Outcomes after Two Years of Exposure for High School Students (Cohort 2 only)

Outcome	Treatment Group		Comparison Group		Impact	Standard Error	p-Value
	Sample Size (schools)	Adjusted Mean	Sample Size (schools)	Unadjusted Mean			
AP pass rate	2	0.07	21	0.04	0.03	0.03	0.298
California College Prepared rate	2	0.27	20	0.29	-0.01	0.10	0.907

AP=Advanced Placement.

The impacts of the program after two years of exposure were also examined for four subgroups in both Cohorts 1 and 2 (Exhibit E-4). The impacts are not significant for most outcomes and subgroups, with the exception of a significant positive effect on science achievement for MLL students (*p*<.05).

Exhibit E-4. Impact of PreK–12 STEM Pathway on Academic Achievement Outcomes After Two Years of Exposure for Selected Subgroups

Outcome	Treatment Sample Size		Comparison Sample Size		Adjusted Treatment Group Mean	Unadjusted Comparison Group Mean	Impact	Standard Error	p-Value
	Schools	School-Grades	Schools	School-Grades					
MLL Students									
ELA	12	41	33	88	-0.39	-0.38	-0.01	0.07	0.851
Math	12	41	34	93	-0.36	-0.36	0.00	0.07	0.956
Science	12	15	33	38	-0.27	-0.43	0.16*	0.07	0.041
Boys									
ELA	12	41	67	191	-0.45	-0.36	-0.09	0.06	0.119
Math	12	41	39	106	-0.29	-0.23	-0.06	0.06	0.301
Science	12	15	32	37	-0.22	-0.25	0.03	0.07	0.662
Girls									
ELA	12	41	40	116	-0.21	-0.14	-0.07	0.06	0.185
Math	12	41	46	127	-0.34	-0.31	-0.03	0.05	0.599
Science	12	15	37	43	-0.16	-0.11	-0.05	0.06	0.460
Hispanic Students									
ELA	12	41	51	134	-0.38	-0.33	-0.06	0.05	0.279
Math	12	41	32	89	-0.50	-0.39	-0.10	0.05	0.054
Science	12	15	38	45	-0.37	-0.30	-0.07	0.07	0.303

ELA=English language arts. MLL=multilingual learner.
* Result is statistically significant at p<0.05.

References

Anwar, S., Menekse, M., Guzey, S.S., & Bryan, L.A. (2022). The effectiveness of an integrated STEM curriculum unit on middle school students' life science learning. *Journal of Research in Science Teaching, 59*(7), 1204–1234. https://doi.org/10.1002/tea.21756

Basche, A., Genareo, V., Leshem, G., Kissell, A., & Pauley, J. (2016). Engaging middle school students through locally focused environmental science project-based learning. *Natural Sciences Education, 45*(1), 1–10. https://doi.org/10.4195/nse2016.05.0012

Bender, R., & Lange, S. (2001). Adjusting for multiple testing—when and how? *Journal of clinical epidemiology*, *54*(4), 343-349. http://www.rbsd.de/PDF/multiple.pdf

Buenrostro, M. (2024). *The State of English Learners in California Public Schools.* Californians Together. Retrieved from: https://californianstogether.org/wp-content/uploads/2024/05/The-State-of-English-learners-in-California-Schools-5.30.24-digital-.pdf

California Department of Education. (2019). *Understanding the college/career readiness measure: Prepared.* https://www.cde.ca.gov/ta/ac/cm/documents/understandcciprepared.pdf

California State Board of Education (2022, April 27). Computer Science Standards Guidelines. https://www.cde.ca.gov/be/st/ss/compsciguidelines.asp

California Department of Education. (2023a). *College/Career Indicator (CCI): Measures of College Readiness*. https://www.cde.ca.gov/ta/ac/cm/documents/ccicollege.pdf

California Department of Education. (2023b). College/Career Indicator downloadable data files. https://www.cde.ca.gov/ta/ac/cm/ccidatafiles.asp

California Department of Education. (2024a). Test Results for California's Assessments. https://caaspp-elpac.ets.org/caaspp/

California Department of Education. (2024b). CAASPP description—*CalEdFacts*. https://www.cde.ca.gov/ta/tg/ai/cefcaaspp.asp

California Department of Education. (2024c). Courses required for California public university. https://www.cde.ca.gov/ci/gs/hs/hsgrtable.asp

California Department of Education. (n.d.) 2021–22 Fixed-length* ASCII research files record definitions. https://caaspp-elpac.ets.org/caaspp/ResearchFileFormatSB?ps=true&lstTestYear=2022&lstTestType=B

California STEM Network. (n.d.). https://csn-search.childrennow.org/

Cantrell, P., Pekcan, G., Itani, A., & Velasquez-Bryant, N. (2006). The effects of engineering modules on student learning in middle school science classrooms. *Journal of Engineering Education, 95*(4), 301–309. https://doi.org/10.1002/j.2168–9830.2006.tb00905.x

Capers, K. J., & Morales C. (2024). *The Academic Outcomes of English Learners Impacted by the COVID-19 Pandemic*. Georgia Policy Labs. https://gpl.gsu.edu/publications/academic-outcomes-of-english-learners-impacted-by-the-covid-19-pandemic/

Caswell, L. (2021). #6420.1v2 PreK–12 STEM Pathway (Early 11). Registry of Efficacy and Effectiveness Studies. https://sreereg.icpsr.umich.edu/sreereg/

Chen, S. Y., Feng, Z., & Yi, X. (2017). A general introduction to adjustment for multiple comparisons. Journal of thoracic disease, 9(6), 1725. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5506159/

College Board. (n.d.). AP score scale table. https://apstudents.collegeboard.org/about-ap-scores/ap-score-scale-table

Columbia University Irving Institute of Medicine. (n.d.). Confirmatory versus exploratory research https://www.irvinginstitute.columbia.edu/confirmatory-versus-exploratory-research#:~:text=In%20a%20study%2C%20investigators%20develop,existing%20hypothesis%20and%20draw%20inferences.

Cunningham, C.M., Lachapelle, C.P., Brennan, R., Kelly, G.J., Tunis, C.S.A., & Gentry, C.A. (2020). The impact of engineering curriculum design principles on elementary students' engineering and science learning. *Journal of Research in Science Teaching*, *57*(3), 423–453. https://doi.org/10.1002/tea.21601

Esquivel, P. (2021, Apr. 4). In California, a million English learners are at risk of intractable education loss. *Los Angeles Times*. https://www.latimes.com/california/story/2021-04-04/how-covid-distance-learning-hurt-california-english-learners

ETS. (2023a). *California assessment of student performance and progress: California Science Test 2020–2021 technical report.* California Department of Education Assessment Development & Administration Division. https://www.cde.ca.gov/ta/TG/ca/documents/casttechrpt21.docx

ETS. (2023b). *California assessment of student performance and progress: Smarter Balanced Summative Assessment 2020–2021 technical report.* California Department of Education Assessment Development & Administration Division. https://www.cde.ca.gov/ta/tg/ca/documents/caaspptechrpt21.docx

Fayer, S., Lacey, A., & Watson, A. (2017). STEM occupations: Past, present, and future. U.S. Bureau of Labor Statistics. https://www.bls.gov/spotlight/2017/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future/pdf/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future.pdf

Feller, A., & Stuart, E. A. (2021). Challenges with evaluating education policy using panel data during and after the COVID-19 pandemic. *Journal of Research on Educational Effectiveness*, *14*(3), 668-675.

Fensterwald, J., & Willis, D.J. (2022, Oct. 24). 2022 California standardized test results wipe out years of steady progress: Smarter Balanced scores dropped in math and English language arts. *EdSource.* https://edsource.org/2022/2022-california-standardized-test-results-wipe-out-years-of-steady-progress/680179#:~:text=California%20skipped%20Smarter%20Balanced%20statewide,test%20optional%20in%202020%2D21

Furner, J.M., & Kumar, D.D. (2007). The mathematics and science integration argument: A stand for teacher education. *Eurasia Journal of Mathematics, Science & Technology, 3*(3), 185–189. https://doi.org/10.12973/ejmste/75397

Guzey, S.S., Harwell, M., Moreno, M., Peralta, Y., & Moore, T. (2017) The impact of design-based STEM integration curricula on student achievement in engineering, science, and mathematics. *Journal of Science Education and Technology, 26*(2), 207–222. https://doi.org/10.1007/s10956-016-9673-x

Guzey, S.S., Moore, T.J., Harwell, M., & Moreno, M. (2016). STEM integration in middle school life science: Student learning and attitudes. *Journal of Science Education and Technology, 25*(4), 550–560. https://doi.org/10.1007/s10956-016-9612-x

Hallberg, K., Williams, R., Swanlund, A., & Eno, J. (2018). Short comparative interrupted time series using aggregate school-level data in education research. *Educational Researcher*, *47*(5), 295–306.

Harris, C.J., Penuel, W.R., D'Angelo, C.M., DeBarger, A.H., Gallagher, L.P., Kennedy, C.A., Cheng, B.H., & Krajcik, J.S. (2015). Impact of project-based curriculum materials on student learning in science: Results of a randomized controlled trial. *Journal of Research in Science Teaching, 52*(10), 1362–1385. https://doi.org/10.1002/tea.21263

Huang, B., Jong, M.S-Y., Tu, Y-F., Hwang, G-J., Chai, C.S., & Jiang, M.Y-C. (2022). Trends and exemplary practices of STEM teacher professional development programs in K–12 contexts: A systematic review of empirical studies. *Computers & Education, 189*, 104577. https://doi.org/10.1016/j.compedu.2022.104577

Hurtt, A., Cohen, K., & Reed, S. (2021). *Early pandemic response in California: Identifying the structural and instructional changes in K–12.* Policy Analysis for California Education. https://files.eric.ed.gov/fulltext/ED612524.pdf

Institute of Education Sciences (Ed). (2013). *Common guidelines for education research and development*. https://www.nsf.gov/pubs/2013/nsf13126/nsf13126.pdf

Johnson, S. (2019). Where to start? Inside one California district's approach to redesign STEM education. *EdSource.* https://edsource.org/2019/where-to-start-inside-one-california-districts-approach-to-redesign-stem-education/614227

Konicek-Moran, R., & Keeley, P. (2015). *Teaching for conceptual understanding in science*. Arlington: NSTA Press, National Science Teachers Association.

Kuhfeld, Megan, James Soland, and Karyn Lewis. (2022). Test Score Patterns Across Three COVID-19-impacted School Years. (EdWorkingPaper: 22-521). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/ga82-6v47

National Academies of Sciences, Engineering, and Medicine. (2019). *Science and engineering for grades 6–12: Investigation and design at the center. Consensus study report.* https://nap.nationalacademies.org/catalog/25216/science-and-engineering-for-grades-6–12-investigation-and-design

National Center for Education Statistics. (n.d.). ACS 2018-2022 Profile. https://nces.ed.gov/programs/edge/TableViewer/acsProfile/2022.

National Center for Education Statistics. (n.d.-a) *Tracy Joint Unified School District*. U.S. Department of Education. https://nces.ed.gov/ccd/districtsearch/district_detail.asp?ID2=0600047

National Center for Education Statistics. (n.d.-b). *Tracy Joint Unified School District: ACS dashboard.* U.S. Department of Education. https://nces.ed.gov/Programs/Edge/ACSDashboard/0600047

National Science Board. (2024). *The state of U.S. science and engineering 2024*. https://ncses.nsf.gov/pubs/nsb20243

Public Policy Institute of California. (2021). *Just the facts: Central Valley.* https://www.ppic.org/wp-content/uploads/content/pubs/jtf/JTF_CentralValleyJTF.pdf

Roehrig, G.H., Dare, E.A., Ellis, J.A., & Ring-Whalen, E. (2021). Beyond the basics: A detailed conceptual framework of integrated STEM. *Disciplinary and Interdisciplinary Science Education Research, 3*(11). https://doi.org/10.1186/s43031-021-00041-y

San Joaquin County. (n.d.). *Industries*. https://www.sjgov.org/business/industries

Schochet, Peter Z. (2008). *Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations (NCEE 2008-4018)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/pdf/20084018.pdf

Somers, M.A., Zhu, P., Jacob, R., & Bloom, H. (2013). *The validity and precision of the comparative interrupted time series design and the difference-in-difference design in educational evaluation.* MDRC. https://www.mdrc.org/work/publications/validity-and-precision-comparative-interrupted-time-series-design-and-difference

St. Clair, T., Cook, T.D., & Hallberg, K. (2014). Examining the internal validity and statistical precision of the comparative interrupted time series design by comparison with a randomized experiment. *American Journal of Evaluation*, *35*(3), 311–327.

U.S. Census Bureau. (n.d.). *San Joaquin County, California*. U.S. Department of Commerce. https://data.census.gov/profile/San_Joaquin_County,_California?g=050XX00US06077

U.S. Department of Education. (2022). *What Works Clearinghouse procedures and standards handbook, Version 5.0.* https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5.0-0-508.pdf

U.S. Department of Education. (2023). *What Works Clearinghouse study review protocol, Version 5.0.* https://ies.ed.gov/ncee/wwc/Document/1297

Thomas, J. (2023). *Girls Who Code "Altamont Connection" brings students from Livermore and Tracy to the Lab.* LLNL. https://www.llnl.gov/article/49806/girls-who-code-altamont-connection-brings-students-livermore-tracy-lab

Wendell, K.B., & Rogers, C. (2013). Engineering design-based science, science content performance, and science attitudes in elementary school. *Journal of Engineering Education, 102*(4), 513–540. https://doi.org/10.1002/jee.20026

Westfall, P. H., Tobias, R. D., & Wolfinger, R. D. (2011). Multiple comparisons and multiple tests using SAS. SAS Institute.